2010

In Vivo Labeling Of A Model β-Clam Protein With A Fluorescent Amino Acid

Mangayarkarasi Periasamy
University of Massachusetts Amherst

IN VIVO LABELING OF A MODEL β-CLAM PROTEIN WITH A FLUORESCENT AMINO ACID

A Thesis Presented

by

MANGAYARKARASI PERIASAMY

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCEINCE

September 2010

Molecular and Cellular Biology

IN VIVO LABELING OF A MODEL β-CLAM PROTEIN WITH A FLUORESCENT AMINO ACID

A Thesis Presented

by

MANGAYARKARASI PERIASAMY

Approved as to style and content by:

_____
Lila M. Gierasch, Chair

_____
Daniel N. Hebert, Member

_____
Matthew Holden, Member

_____
John M. Lopes, Interim Director
Molecular and Cellular Biology Program

ACKNOWLEDGEMENTS

I thank **Prof. Lila Gierasch** for providing me the opportunity to work with her research group. I thank everyone of the Gierasch lab for helping me all along. I would like to specifically thank **Dr. Beena Krishnan** for her utmost support throughout this work, **Dr. Jenny Maki**, for help and suggestions, **Dr. Harekrushna Sahoo** and **Dr. Annie Marcelino** for assisting me in finding suitable positions for labeling, **Dr. Qinghua Wang** for helping me with the NMR, **Dr. Ivan Budyak, Dr. Eugenia Clerico** for assisting me at various times and also **Dr. Anne Gershenson, Dr. Jiang Hong, Rob Smock, Mylene, Kristine, Dr. Santosh Kumar, Dr.Anastasia Zhuravleva, David Paquette, Ervin Pejo** and **Amanda Clouser** for their support.

I would like to thank **Stephen Eyles** for the mass spectrometery measurements. I also thank my chemistry graduate student friends for their valuable suggestions regarding the synthesis procedure. I specially thank **Ellen Kalt** for all her help and care throughout.

I dearly thank my lovely parents, my little brother and all my family for being with me through all my times and encouraging me for a better tomorrow. I hope for the best in life with all their blessings.

ABSTRACT

IN VIVO LABELING OF A MODEL β-CLAM PROTEIN WITH A FLUORESCENT AMINO ACID

SEPTEMBER 2010

MANGAYARKARASI PERIASAMY,

B.S., NIZAM COLLEGE INDIA

M.S., OSMANIA UNIVERSITY INDIA

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Lila M Gierasch

Proteins can be labeled with different tags to enable their structural and functional investigations. In addition, labeling proteins at specific sites helps in studying the conformational dynamics of these molecules. A plethora of methods is available to facilitate labeling, choice of which largely depends on the requirements and the anticipated end results. In general, the various labeling methods can be classified into four different classes based on the stage at which labeling is performed, namely post translational labeling, non-ribosomal synthesis, *in vitro* translation and *in vivo* translation. Interestingly all these techniques use different unnatural amino acids for this purpose.

Protein folding is one among the many applications that requires tailoring proteins with special molecules or labels for deducing the folding pathway. Understanding the protein folding problem is a key for answering questions concerning protein behavior and thus, will provide strategies to solve protein misfolding diseases. Protein folding is one among the unsolved problems in biology and in particular understanding the *in vivo* behavior of proteins in the complex cytoplasm environment with a cellular density of approximately 350 to 400 mg/ml is more critical. It is evident that there is a difference in the behavior and folding of proteins *in vivo* and *in vitro* and to deduce more insights in this aspect the

protein of interest is to be labeled with a sensitive probe. The *in vivo* translation method offers a good method of choice for labeling the protein at a specific position and monitoring its behavior.

To study the ultimate goals of acquiring knowledge of the *in vivo* behavior and folding characteristics of proteins, the first step of establishing an efficient labeling technique is quintessential and as a starting step, this project aims to label a β-clam protein, cellular retinoic acid binding protein I (CRABP I) a 136 amino acid protein, with a sensitive unnatural fluorescent amino acid probe *in vivo* in *E. coli* cells.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1
INTRODUCTION

1.1 Protein labeling strategies

Designing and tailoring proteins are major tools facilitating molecular studies of various types. Apart from being tools in enhancing protein investigations, these designer proteins will make their way into many other applications too (Dougherty, 2000; Wang and Schultz, 2004). This is promising especially for protein folding investigations. Labeling proteins can be essentially performed by post-translational modification, non-ribosomal synthesis and *in vitro* and *in vivo* translation methods using non natural amino acids. Non natural amino acids serve as spin labels, fluorescent amino acids, cross-linking agents and other spectroscopic and mechanistic probes (Dougherty, 2000; Hendrickson et al., 2004). In this process, scientists have managed to extend the 64 triplet genetic code allowing for decoding the 'amber' stop codon for a non natural amino acid. This library of available unnatural amino acids is presently being extended. These are amino acids designed to serve a purpose which will exclusively lead to production of labeled proteins with the desired unnatural amino acid at a specific position that will eventually serve as a tag for molecular protein studies.

The available methods for incorporation of unnatural amino acids include: use of amino acid auxotrophs ('residue specific' incorporation of non-canonical amino acid) (Kiick et al., 2001), *in vitro* translation with chemically altered aminoacylated tRNAs, SELEX of ribozymes, solid phase peptide synthesis (SPPS), cyanogens bromide cleavage and peptide religation, import of aminoacylated amino acids into eukaryotic cells, unnatural amino acid incorporation using orthogonal aminoacyl synthetase and tRNA pairs and chemical modification of amino acids (Baslé et al., 2010). Among these the orthogonal

synthetase and tRNA pair can be specifically used complementary to nonsense codon, four base or five base codons (Kiick et al., 2001).

1.2 Method of choice for labeling

The phenomenon of nonsense suppression has been known and worked upon from the 1960's. However, the first attempts to introduce unnatural amino acids at specific sites in a protein were accomplished in a cell free translation system by the Peter Schultz group (Noren et al., 1989). They employed chemically acylated tRNAs with the unnatural amino acids for this purpose. The *in vivo* approach of the same technique was started with the aim to enhance the value of the method in cells and avoid the laborious biochemical procedure involved in the former.

The beauty of the method is the use of the cell's own machinery to translate the labeled protein unlike in the *in vitro* techniques. This technique when compared to the *in vitro* translation labeling can be viewed being cost effective and also efficient (Ohno et al., 1998). This method has paved the way for using small chemical molecules with special properties to label proteins, unlike the available labels that are comparatively larger in size and toxic to cells. In comparison with the chemical aminoacylation of the tRNAs *in vitro* and subsequent microinjection into eukaryotic cells this method impressively exploits the cells machinery for doing so. This technique allows for the desired protein modification under native conditions in the intracellular environment.

When compared to the 'residue specific' incorporation of non-canonical amino acid in which the sense codons coding for a particular natural amino acid are reassigned with the unnatural or non-canonical amino acid by altering the amino acid pool (Link and Tirrell, 2005), the *in vivo* nonsense codon suppression technique is highly specific in terms of

position and not residue. Unlike 'residue specific' incorporation, this process lacks the editing step consequent to misacylation of tRNAs with unnatural amino acids in the ribosome (Effraim et al., 2009). There are often problems of introducing analogues nonselectively at the locations for the residue specific incorporation (Furter, 1998). However following steps might be effected which could be a plausible reason for the low level of expression. These factors are still unknown to large extent. In addition the nonsense suppression technique has been applied in mammalian cells, yeast, *E. coli* and also other bacterial genera like *Mycobacterium* (Wang et al., 2010).

Nonsense suppression mediated labeling is facilitated by the presence of the unnatural amino acid in the medium of the *E. coli* culture, specific tRNAs that can recognize the amber stop codon and bind specifically to the unnatural amino acid in the amino acid binding pocket and the specific aminoacyl tRNA synthetase that recognizes the specific loaded tRNA and the unnatural amino acid.

1.3 Nonsense suppression technique

There are three mRNA codons recognized as stop signals for protein translation, viz. amber (UAG), opal (UGA) and ochre (UAA). Among these there is a preference for use of a particular stop codon more frequently in particular living systems for example, *E. coli* protein genes carry the ochre stop codon as the stop signal whereas eukaryotic systems use the amber codon for the same purpose. This fact is exploited to introduce an unnatural amino acid apart from the twenty natural amino acids at the infrequently used stop codon in *E. coli*.

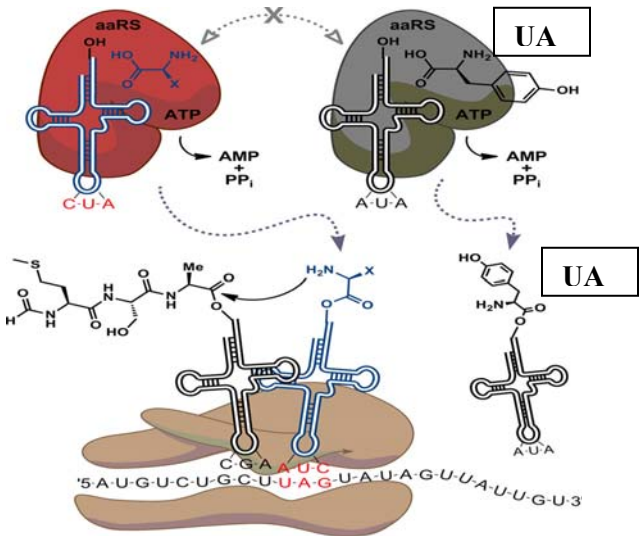


Fig 1.1 Schematic of the working mechanism of the nonsense suppression technique. The ribosomes are shown in red, grey and brown. The various ingredients of the translation process are depicted, viz. tRNA, mRNA, unnatural amino acid (UA) acylated to the respective tRNA (in blue)(Wang and Schultz,2004).

For any amino acid to be incorporated at a particular position in a polypeptide it has to be recognized by a specific tRNA to which it should be acylated by a specific aminoacyl tRNA synthetase with specific recognition for the former two. The tRNA carries the complementary anticodon to the particular sequence on the mRNA. Hence the fidelity of the reaction is mainly attributed to the recognition of the entities and hence the specificity.

These are taken advantage of to facilitate labeling of a protein during the process of translation. This technique makes it possible to incorporate an unnatural amino acid of choice at a position in a protein encoded by the amber (UAG) stop codon. In the presence of a modified tRNA that recognizes the UAG codon at the particular position and the specific aminoacyl tRNA synthetase the unnatural amino acid of choice is incorporated in the protein as shown in Fig 1.1. To avoid the ability of the tRNA to get acylated by any of the cell's own synthetases the system is so designed that both the tRNA and the

synthetases are 'orthogonal' to the system used for protein expression and has been shown not to be involved in the cells native functions. Since the process takes place *in vivo*, the other requirements for the translation process to take place are taken care of by the cells.

The ingredients that function in this incorporation method are mutant orthogonal tRNA, mutant orthogonal aminoacyl tRNA synthetases and the unnatural amino acid. Each of these ingredients has certain properties aiding for their function and each of these is being discussed ahead.

1.4 Ingredients for nonsense suppression technique

- Suppressor tRNA

The suppressor tRNA used must meet two criteria: 1) it must efficiently insert the desired amino acid in response to the UAG codon, and 2) it must not be acylated by any of the *E. coli*'s aminoacyl tRNA synthetases present *in vivo*. These two factors are necessary to obtain the desired protein to be labeled and not to obtain any mislabeled protein product. The amber suppressor tRNAs in our experiments are derived from *Methanococcus jannaschii*, a thermophilic archaeabacterium. The tRNA is an $MjtRNA^{Tyr}_{CUA}$ originally destined to carry a tyrosine molecule. It has been observed that the same tRNA derived from *S. cerevisiae* and *H. sapiens* is not orthogonal in *E. coli* (Wang and Schultz, 2004). The *M. jannaschii* tRNA was hence chosen. The Schultz group has mentioned that although the tRNA is well recognized by the corresponding aminoacyl MjtRNA synthetase it can also at low levels be aminoacylated by the *E. coli* aminoacyl tRNA synthetases.

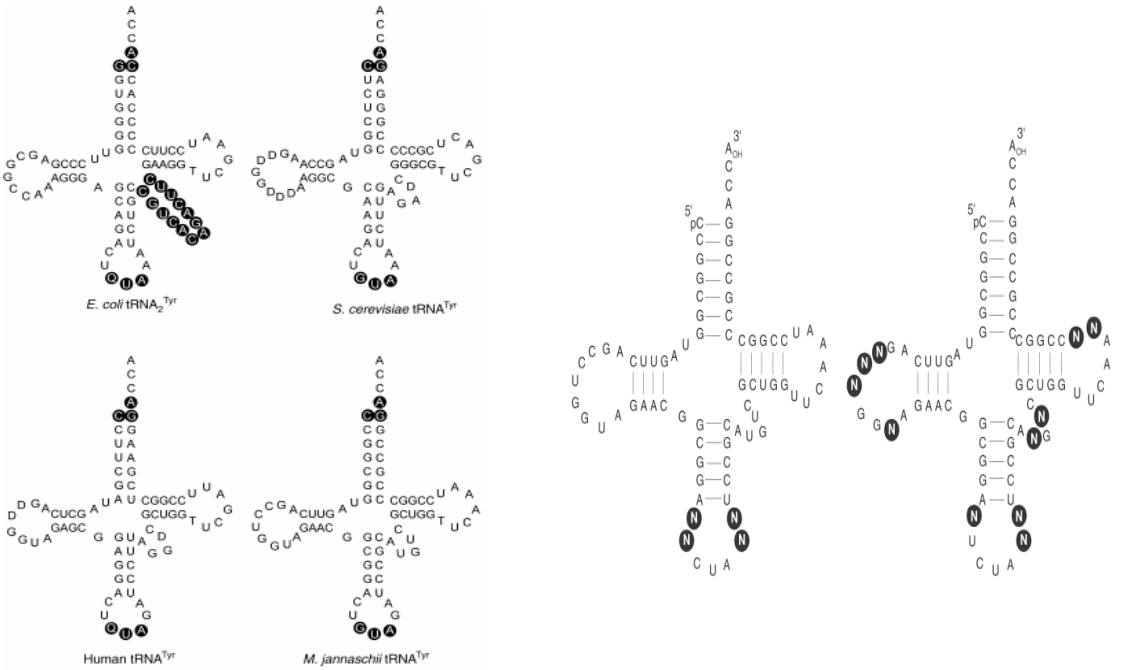


Fig. 1.2 Structure of tRNAs A.Comparison of the structure of the tRNAs from different organisms. B.Anticodon-loop tRNA library (left) and all-loop tRNA library (right) derived from *M. jannaschii* tRNATyr CUA. Randomly mutated nucleotides (N) are shaded in black.

To overcome this, the group had made positive and negative selections with a mutant tRNA library in the absence and presence of the cognate aminoacyl tRNA synthetases in *E. coli*. The amber codon in barnase gene was used for selection of clones that could only survive when provided with the orthogonal tRNA. The next selection was done for the cognate aminoacyl tRNA synthetases (Wang and Schultz, 2001).

It is known that the concentration of suppressor tRNA in the cell may cause detrimental effects to the cell. Hence, the level of expression of the tRNA is also to be regulated (Miller, 1991).

- Aminoacyl-tRNA synthetases

Mutant orthogonal aminoacyl tRNA synthetases are used to suppress the amber codon in *E. coli* and incorporate the unnatural amino acid in response to the amber codon. Methyl-O-tyrosine was one among the first candidates to be used in the field of nonsense

suppression being recognized by orthogonal MjtRNA and aminoacyl tRNA synthetase pair specially mutated and selected for this purpose (Wang and Schultz, 2004).

Tyrosyl tRNA synthetase has no proofreading capacity and hence is chosen against Glutaminyl tRNA synthetase earlier studied. It has also been observed that the synthetases from archaea can be efficiently expressed in *E. coli* in the active form as opposed to the eukaryotic counterparts. The aminoacyl MjTyrtRNA synthetase discriminates tRNAs with C1:G72 (an eukaryotic recognition element) from those with G1:C72 (prokaryotic recognition element) and it does not aminoacylate *E. coli* tRNA. The aminoacyl MjTyrtRNA synthetase is devoid of the C-terminal binding domain for the anticodon loop of its $tRNA^{Tyr}$. This shows that the recognition elements of their aminoacyl tRNA synthetase is far different from the *E. coli's* and provides the basis for orthogonality. Another factor that it facilitates is that the mutation in the anticodon loop of the tRNA to recognize amber codon would not interfere and affect the recognition and binding. Structure-based mutagenesis and a similar two-step selection strategy are used to alter the specificity of the heterologous aminoacyl tRNA synthetase so that it uniquely recognizes the unnatural amino acid (UA) of interest.

The fidelity of translation of a polypeptide is determined by the availability of the amino acyl-tRNA composed of the cognate amino acid:tRNA pairs and the accurate selection of the aminoacyl-tRNAs on the ribosome (Ling et al., 2009). Amino acids that differ by for example a single methyl group are often not that well differentiated by the aminoacyl tRNA synthetases and hence contributing to the translation error rate. However, the cell is equipped with the necessary editing machinery to cope up with such situations. This editing machinery is deceived in this process of nonsense suppression. Nevertheless,

there is a percentage of loss during this process and hence lowering the yield of the protein being formed. The engineered aminoacyl MjTyrtRNA synthetase structure has been obtained by x-ray crystallography (Wang and Schultz, 2001).

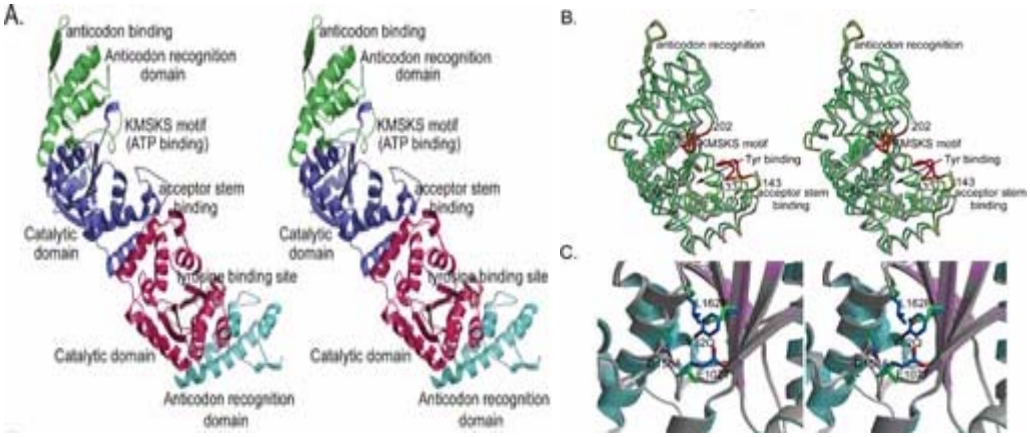


Fig.1.3 Crystal structure of apo wild-type *M. jannaschii* TyrRS. (*A*) Overall structure of apo *M. jannaschii* TyrRS homodimer. The N-terminal catalytic domain (blue and red) is connected via the signature KMSKS loop to the C-terminal anticodon recognition domain (green and cyan). (*B*) Stereo view of the superimposition of wild-type *M. jannaschii* TyrRS and OMeTyr-specific mutant synthetase to illustrate conformational differences. Apo mutant *M. jannaschii* TyrRS is colored by its rms deviation (Cα) from the apo wild-type TyrRS structure (gray) with a gradient ranging from low (0.2 Å, green) via medium (1.8 Å, yellow) to high (>4.0 Å, red). (*C*) Stereo view of OMeTyr-specific mutant synthetase superimposed with wild-type apo *M. jannaschii* TyrRS (gray) to highlight mutated residues. The mutant enzyme is colored with cyan helices and pink β-strands. The four residues subjected to mutation are represented by ball-and-stick with their oxygen atoms colored red. Bonds/carbon atoms are colored green in the mutant enzyme but blue in wild-type (Zhang et al., 2005).

The aminoacyl MjTyrtRNA synthetase is 18 % identical to that of *Bacillus stearothermophilus*, 22 % identical to that of *Thermus thermophilus* and 37 % identical to the human TyrRS catalytic domain. While the overall fold of the apo aminoacyl MjTyrtRNA synthetase is identical to its counterpart with tRNA and tyrosine with an overall rmsd (Cα) of 1.5 Å, the active-site mutations result in hydrogen bonding and steric interactions favoring binding of Methyl-O-tyrosine over L-tyrosine. It has been found that the aminoacyl MjtRNA synthetases have higher degree of plasticity when compared to the other counterparts that enables their ability to be used for nonsense

suppression (Turner et al., 2006). The tRNA and synthetase pair was first discovered for suppression of the amber codon using the endogeneous amino acid pool (Wang and Schultz, 2001).

1.5 Essential elements for the feasibility of the nonsense suppression technique

1.5.1 Overexpression of aminoacyl-tRNA synthetases

The activity of this enzyme is a critical controlling factor in determining the success rate of the incorporation of the unnatural amino acid at the specific position.

The different promoter elements controlling the synthetase enzyme expression developed by the Schultz group have recently been tested for the incorporation efficiency using amber mutant GFP (Young et al., 2010). It is evident from these results that the gene for synthetase enzyme when under control of the *E. coli* constitutive *glnRS* promoter yields lesser GFP product than under control of *araBAD* promoter. However contrasting results have been observed by these workers in other systems. The synthetase gene for the fluorescent amino acid and the methyl tyrosine incorporation are under the control of the *E. coli* constitutive *glnRS* promoter. However the expression level of the product in case of the latter is higher than with the fluorescent amino acid showing that this expression level (definitely lower than the wild type expression) of synthetases is not the critical factor.

1.5.2 Aminoacyl-tRNA synthetase activity

The unnatural amino acid is activated by binding to the binding pocket in the synthetase enzyme. The editing activity by which misincorporation is avoided by the natural system is inactivated in this case hence allowing for the incorporation of the unnatural amino acid (Hendrickson et al., 2004).

1.5.3 Amber stop codon

The UAG amber stop codon is used for the assignment of an unnatural amino acid since this codon is more easily suppressed by a tRNA bearing CUA anticodon competing with the release factor 1 (RF1) in *E. coli* (Hayashi et al., 2010). The amber codon is a rarely used stop codon in *E. coli* i.e 93 % of *E. coli* genes end with TAA or TGA and not TAG (Young and Schultz, 2010). The criterion for the nonsense suppression technique is that the unique codon must be recognized by the new tRNA but not by any endogenous tRNAs hence the amber codon will serve this purpose well. It has been shown that expressing proteins with rare codons used by bacteria leads to long pauses by the ribosome leading to differences in folding and insolubility of the proteins. However if provided with the required tRNAs this will be combated for. Decreasing the temperature while the bacteria are expressing the protein leads to increasing the solubility of the protein and lowers the overall protein synthesis preventing saturation of the cellular folding machinery with the over expressed recombinant proteins (Rosano and Ceccarelli, 2009). Another aspect to keep in mind is the context of the amber codon. UAG followed with A are suppressed efficiently, if followed by G are moderately suppressed and those followed by U and C are said to be poorly suppressed. The CUX codon after the UAG however is said to be efficient. The AUX is the most efficient. In our case the UAG codon is followed by ACT. Since there is an A post UAG it is supposed to be efficient (Miller, 1991).

1.5.4 Unnatural amino acids

The 21st amino acid selenocysteine is incorporated in response to UGA stop codon which forms the basis for expanding the genetic code in *E. coli* for labeling proteins (Chambers et al., 1986). Amino acids with different chemical handles, fluorescent probes, redox active groups or heavy atoms offer a great probe tool box for manipulating protein function *in vitro* and *in vivo* (Chin et al., 2002; Deiters et al., 2005; Liu et al., 1997; Fleissner et al., 2009; Hartman et al., 2007). One of the major requirements of the procedure to be successful is the ability for the amino acid to be transported to the cytoplasm of the cell from the medium. The unnatural amino acid should be metabolically stable, should be available for the cells machinery, it must be tolerated by EF-Tu and the ribosome, but it must not be a substrate for any endogenous aaRSs. Most unnatural amino acidss added to the media are taken up by both prokaryotic and eukaryotic cells (exceptions include highly charged amino acids, which can be modified as metabolically labile derivatives or incorporated into dipeptides to increase permeability) (Zhang et al., 2005). Some unnatural amino acids that have already been applied for labeling proteins are 2-amino-3-(4-(trifluoromethoxy) phenyl) propanoic acid (OCF3Phe), 13C/15N-labeled *p*-methoxyphenylalanine (OMePhe), and ^{15}N-labeled o-nitrobenzyl-tyrosine (oNBTyr) (Cellitti et al., 2008), p-azido-L-phenylalanine (Chen et al., 2008), benzyl phenylalanine (Ryu and Schultz, 2006; Chin et al., 2002), *p*-acetyl-L-phenylalanine (Fleissner et al., 2009).

L-(7-hydroxycoumarin-4-yl) ethyl glycine (HCEG) was chosen for the incorporation into CRABPI on account of its high quantum yield, large Stoke's shift, small size and sensitivity to pH (Adamczyk et al., 1997). The other amino acid that is chosen to support

the labeling studies is Methyl-O-tyrosine. The methyl-O-tyrosine is a tyrosine analog with the para hydroxyl group substituted with a methoxy group (Wang and Schultz, 2001).

Transport of the unnatural amino acid into the *E. coli* cells takes place passively. Polar molecules are transported across membranes through transmembrane channels. It is known that unlike small molecules like ions large molecules like glucose and amino acids are transported through the membranes by special carriers like permeases. This process is energy independent.

1.5.5 Efficient systems for nonsense suppression expression

Considering the pitfalls especially low level expression of labeled protein, certain amount of tweezing of the plasmid and respective promoters have been made to allow for higher level expressions of the tRNA, synthetase and eventually in the process the labeled protein (Ryu and Schultz, 2006). Plasmid systems with multiple copies (polycistronic) of the tRNA gene under the *proK* promoter are designed and the efficiency of protein expression has been claimed to be increased to 40 mg/l from 2 mg/l (Ryu and Schultz, 2006).

The nonsense suppression technique is hence a reasonably good method of choice for labeling proteins *in vivo* and offers promising strategies if all the parameters explained before are satisfied.

1.6 Labeling a β-clamp protein

We are interested to study the *in vivo* dynamics of CRABP I in the *E. coli* cytoplasm for which we have chosen to label the protein with a fluorescent unnatural amino acid *in vivo*

using the nonsense suppression technique. A general schematic of the goal is shown in Fig 1.4.

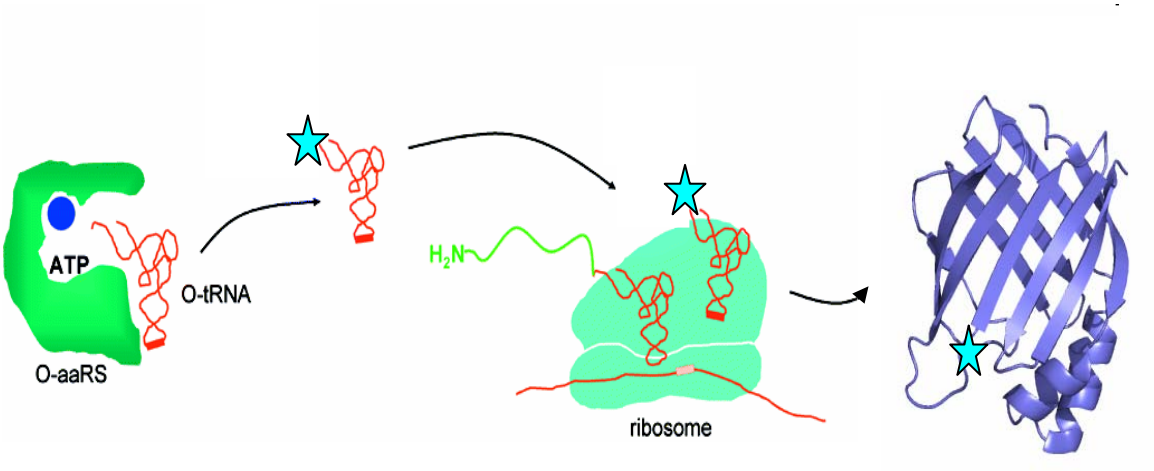


Fig 1.4 Schematic representation of labeling of CRABP I with the fluorescent amino acid using the nonsense suppression technique (Wang et al., 2006)

CRABP I is a 136 amino acid long model β-clam protein whose *in vitro* folding mechanism has been well studied in the lab. However, very less is known about the *in vivo* folding mechanism of this protein which is expected to be very different from the *in vitro* observations on account of the macromolecular crowding in the cytoplasmic context. The average cytoplasmic density in *E. coli* is estimated to be around 350-400 mg/ml. It can be predicted that this high density will induce variability in the in vivo folding mechanism of the protein which would like to be investigated using in vivo tools. Fluorescence labeling offers a good scope of learning this aspect in addition to being a sensitive technique. *In vivo* fluorescence resonance transfer (FRET) experiments will be performed using compatible dyes at different calculated positions to study the *in vivo* protein dynamics and as a first step labeling of the protein *in vivo* at a specific position with a comparatively small fluorescent dye has been done as a first step to accomplish these studies.

CHAPTER 2
MATERIALS & METHODS

2.1 Choice of the unnatural amino acid

L-(7-hydroxycoumarin-4-yl) ethyl glycine was chosen on account of its high fluorescence quantum yield of 0.63, large Stoke's shift of approximately 100 nm, small size of 264 Da and sensitivity to pH. It was synthesized following the Pechman condensation reaction and purified. Incorporation in response to the amber codon TAG by the orthogonal pair of *Methanococcus jannaschiii* $tRNA^{TYR}_{CUA}$ and *Mj*TyrRS was performed. Apart from these properties, the chosen fluorescent amino acid has an excitation maximum at 350 nm and emission maximum at 450 nm hence it forms good FRET pairs with FlAsH (excitation maximum of 510 nm and emission maximum of 530 nm; L-3-(2- Naphthyl) alanine(excitation maximum of 285 nm and emission maximum of 335 nm; BODIPY-FL (excitation maximum of 488 nm and emission maximum of 510 nm) and tryptophan (excitation maximum of 295 nm and emission maximum of 350 nm).

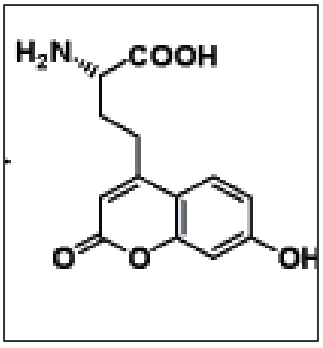


Fig 2.1 Structure of L-(7-hydroxycoumarin-4-yl) ethyl glycine

Site directed mutagenesis experiments were designed to introduce the TAG codon into the CRABP I gene at a particular position. Following which cloning to introduce the mutated CRABP I gene into the required plasmid was performed. Co-expression of the stop codon containing CRABP I, *Mj*$tRNA^{TYR}_{CUA}$ and aminoacyl MjTyrtRNA synthetase was performed. Further to optimize the labeled protein expression and learn the working

of the technique another unnatural amino acid system namely Methyl-O-Tyrosine was also used.

Table 1 is a summary of the various unnatural amino acids developed for incorporation in various systems among which many are under investigation. (Wang et al., 2010).

Unnatural amino acid (UAA)	Common name (if applicable)	Organism(s) in which UAA is encoded[a]
1	*p*-Acetylphenylalanine	*E. coli*, yeast, mammalian cells
2	*m*-Acetylphenylalanine	*E. coli*
3		*E. coli*
4	*O*-allyltyrosine	*E. coli*
5	Phenylselenocysteine	*E. coli*
6	*p*-Propargyloxyphenylalanine	*E. coli*, yeast, mammalian cells
7	*p*-Azidophenylalanine	*E. coli*, yeast, mammalian cells
8	*p*-Boronophenylalanine	*E. coli*
9		*E. coli*
10		*E. coli*
11		*E. coli*
12	*O*-methyltyrosine	*E. coli*, yeast, mammalian cells
13	*p*-Aminophenylalanine	*E. coli*
14	*p*-Cyanophenylalanine	*E. coli*
15	*m*-Cyanophenylalanine	*E. coli*
16	*p*-Fluorophenylalanine	*E. coli*
17	*p*-Iodophenylalanine	*E. coli*, yeast, mammalian cells
18	*p*-Bromophenylalanine	*E. coli*
19		*E. coli*
20	*p*-Nitrophenylalanine	*E. coli*
21	L-DOPA	*E. coli*
22	3-Aminotyrosine	*E. coli*
23	3-Iodotyrosine	*E. coli*, yeast, mammalian cells
24	*p*-Isopropylphenylalanine	*E. coli*
25	3-(2-Naphthyl)alanine	*E. coli*
26	Biphenylalanine	*E. coli*
27		Yeast, mammalian cells
28		Yeast, mammalian cells
29		Yeast, mammalian cells
30	Homoglutamine	*E. coli*
31	D-tyrosine	*E. coli*
32	*p*-Hydroxyphenyllactic acid	*E. coli*
33	2-Aminocaprylic acid	Yeast, mammalian cells
34	Bipyridylalanine	*E. coli*
35	HQ-alanine	*E. coli*
36	*p*-Benzoylphenylalanine	*E. coli*, yeast, mammalian cells
37	*o*-Nitrobenzylcysteine	Yeast, mammalian cells
38	*o*-Nitrobenzylserine	Yeast, mammalian cells
39	4,5-Dimethoxy-2-nitrobenzylserine	Yeast, mammalian cells
40	*o*-Nitrobenzyllysine	*E. coli*, yeast, mammalian cells
41	*o*-Nitrobenzyltyrosine	*E. coli*
42	2-Nitrophenylalanine	*E. coli*
43		*E. coli*
44		*E. coli*
45	Dansylalanine	Yeast, mammalian cells
46		*E. coli*
47		Yeast, mammalian cells
48	*p*-Carboxymethylphenylalanine	*E. coli*
49	3-Nitrotyrosine	*E. coli*
50	Sulfotyrosine	*E. coli*
51	Acetyllysine	*E. coli*, yeast, mammalian cells
52	Methylhistidine	Yeast, mammalian cells
53	2-Aminononanoic acid	Yeast, mammalian cells
54	2-Aminodecanoic acid	Yeast, mammalian cells
55		Yeast, mammalian cells
56		Yeast, mammalian cells
57		Yeast, mammalian cells
58		Yeast, mammalian cells
59	Pyrrolysine	*E. coli*, yeast, mammalian cells
60	Cbz-lysine	*E. coli*, yeast, mammalian cells
61		*E. coli*, yeast, mammalian cells
62		*E. coli*, yeast, mammalian cells
63		*E. coli*, yeast, mammalian cells
64		*E. coli*, yeast, mammalian cells
65	Boc-lysine	*E. coli*, yeast, mammalian cells
66		*E. coli*, yeast, mammalian cells
67	Allyloxycarbonyllysine	*E. coli*, yeast, mammalian cells
68		*E. coli*, yeast, mammalian cells
69		*E. coli*, yeast, mammalian cells
70		*E. coli*, yeast, mammalian cells
71		*E. coli*, yeast, mammalian cells

[a]Underlined font, functionality not experimentally demonstrated but based on parent aminoacyl-tRNA sy
[b]References are those pertinent to the original encoding of the UAA.
[c]Abbreviation: UAARS, UAA-specific mutant aminoacyl-tRNA synthetase.

Table 1. List of the various unnatural amino acids incorporated in proteins using the *in vivo* nonsense suppression method (Wang et al., 2010).

- Amber codon in CRABP I:

The Lysine 106 position in CRABP I was chosen for mutagenesis to the amber codon. This position is present in the Ω loop as observed in Fig 2.2. The sequence variation in the Ω loop among various iLBP (intracellular Lipid Binding Protein) members shows low sequence conservation (Kohrer et al., 2003) minimal or no secondary and tertiary structure perturbation, accessible molecular surface, other possible interactions and free accessibility.

- Site of incorporation:

The incorporation of the unnatural amino acid at the specific site has been known to have a relation with the site chosen for incorporation. However, the exact rules governing this factor are not yet formed (Young et al., 2010). Using an earlier position in the gene of interest restricts the observation of truncated products. In the 1998 Ohno et al. (Ohno et al., 1998) have used the 126 position in the β-lactamase gene but only measure the activity of the protein. While in papers later an earlier position facilitating N-terminal sequencing (Furter, 1998) has been used. However, there is also evidence of truncation (Liu et al., 1997).

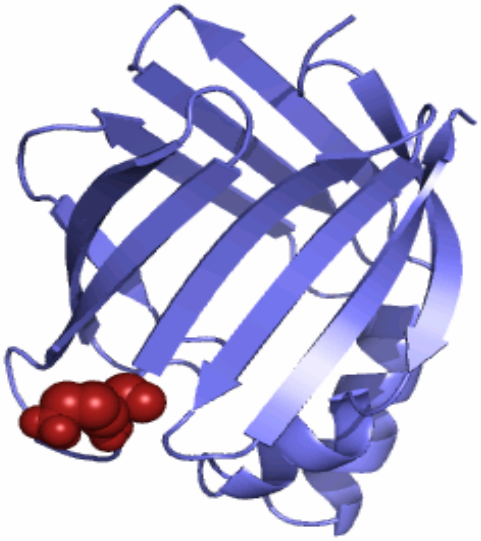


Fig 2.2 Structure of CRABP I. The Lysine-106 position is shown as red spacefill.

2.3 Plasmids and genes

The pBAD/JYAMB-4TAG-Myo and pBK-CouRS-D8 plasmids (Wang et al., 2006) were offered by the Schultz lab. The pBAD plasmid carries the mutant sperm whale myoglobin gene with the amber codon at the Serine 4 position, mutated *Mj*tRNA gene and tetracycline resistance gene as a selective marker.

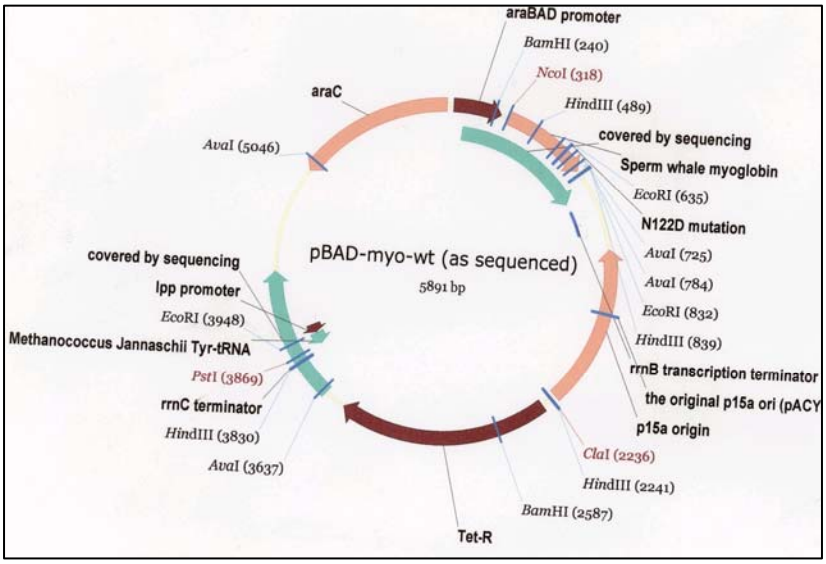


Fig 2.3 Vector map of pBAD-myo-wt plasmid as provided by the Schultz laboratory

The myoglobin gene is under the arabinose promoter and *rrnB* terminator. The *Mj*tRNA gene is with the *lpp* promoter and *rrnC* terminator. The lpp (lipoprotein) promoter is a highly efficient constitutive promoter usually used for *in vivo* RNA expression (Umekage and Kikuchi, 2009). The pBK plasmid carries the aminoacyl MjTyrtRNA synthetase (MjTyrRS) gene in which the residues have mutations as follows; Tyr32 Glu, Leu65 His, Ala67 Gly, His70 Gly, Phe108 Tyr, Gln109 His, Asp158 Gly and Leu162 Gly. The four glycine mutations offer space for accommodation of the unnatural amino acid. Tyr32 and Asp158 which are involved in hydrogen bonding to the native substrate, tyrosine are

mutated to loose activity towards tyrosine. The plasmid has the kanamycin gene as a selection marker.

For the Methyl-O-Tyrosine incorporation the pSup-JYRS plasmid was obtained from the Schultz group. This system has been observed to be more efficient for unnatural amino acid incorporation when compared to the pBK system (Young et al., 2010). The vector carries the JYRS (synthetase), two tricistronic clusters of the tRNA (JYTRN) gene and chloramphenicol resistance gene. The synthetase gene is under the control of a modified *glnS'* promoter. The tRNA genes are under the control of the *proK* promoter.

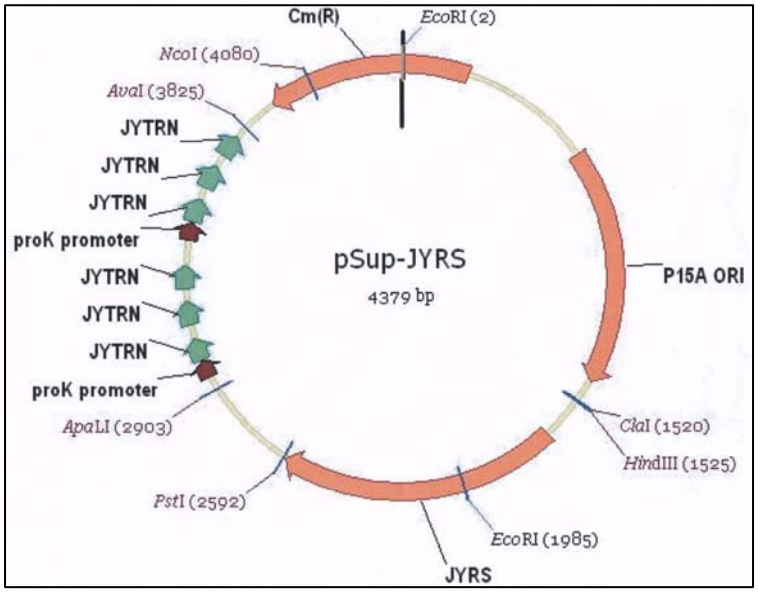


Fig 2.4 Vector map of pSup-JYRS plasmid for the incorporation of O-Methyl-tyrosine (as provided by the Schultz group)

2.4 Comparison of the systems for 7-Hydroxycoumarin-4-ethyl glycine and methyl-O-tyrosine incorporations:

Young et al. have provided comprehensive explanation for the new plasmid system (pSup) with enhanced expression levels of labeled protein with unnatural amino acids. These plasmids are designed to facilitate the gene of interest to be encoded in colE1 or pBR322 origin plasmids. In this case the gene of interest can be under a system of higher

expression like the T7 promoter, unlike pBAD. This facilitates higher expression of the labeled protein. A similar system has been constructed for the fluorescent amino acid labeling recently. However, the expression of the tRNA genes in case of pSup vector under the control of the *prok* promoter has been observed to slow down the growth phenotype. This system has been used for longer induction times of 12-14 Hrs and it has been observed that the growth temperature had no particular effect on the expression of the protein.

2.5 Competent cells:

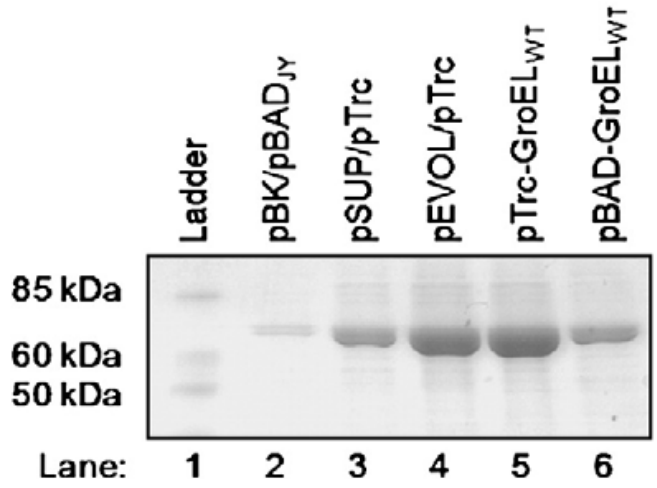


Fig 2.5 Comparison of the protein expression levels using different incorporation systems (From Travis S.Young et al., 2010)

The type of competent cells also plays a role in expression of the various components and the final expression of the labeled product. Expression of GroEL using different vector systems in different competent cells has been performed and it is found that the expression from pSUP is best in BL21(DE3) cells where as expression from pBK is best in DH10B cells (Young et al., 2010). As observed in the Fig 2.5 the expression levels while using the pSup system is higher when compared to that of the pBAD-pBK. Based

on these they have developed a pEVOL plasmid system for the incorporation of the fluorescent amino acid into any protein of interest that can be encoded in a pET vector. This would be a simpler system giving high yields of labeled protein expression for us. The system has been requested from the Schultz lab.

2.6 Culture Medium:

Suppression of protein with amber codon when expressed with synthetase gene has been observed even in the absence of the unnatural amino acid in 2XYT (rich medium). However this background is not observed when expressed in GMML medium (minimal medium). It should be noted that this kind of background incorporation occurs by aminoacylation of the suppressor tRNAs by the evolved synthetase with an endogeneous amino acid (usually Phe, Tyr) in rich media in the absence of the cognate unnatural amino acid. This background incorporation is typically suppressed by the addition of cognate unnatural amino acid to the expression media (Young et al., 2010)**.**

2.7 Factors to consider while using the nonsense suppression technique

Based on literature and results the following can be laid out as basic considerations while using the nonsense suppression technique for labeling proteins *in vivo*;

1. The codons chosen to be mutated to the amber codon are minor/major for the particular amino acid. Very often the neighboring residues are important in the interpretation of the nonsense codon by the ribosome due to wobble effect (Finkel, 2010).

2. While designing the site for UAG mutation, care was taken to prevent the presence of unfavorable amino acids like leucine, arginine and serine codons following the UAG as these are highly degenerate.

The amino acid pool in the *E. coli* cytoplasm was defined by researches much earlier (Britten and McClure, 1962). Researchers usually use M9 minimal medium for labeling purposes as this would define the amino acid pool in *E. coli* (Link and Tirrell, 2005). However, the Schultz group has obtained similar labeled protein yields in the rich 2XYT medium and in the minimal medium M9. Hence, unlike the residue specific labeling wherein the cells are regulated with the amino acids provided with, this method does not require the stringency in that sense. The only major limiting factors regarding the presence of the unnatural amino acid in the total cytoplasmic amino acid pool are the metabolic stability of the unnatural amino acid and the bioavailability of the amino acid to the cells.

The kinetics of adenylate formation with ATP catalyzed by the mutant TYRS was tested by the Schultz group *in vitro* and was found that the K_{cat}/K_m of the mutant TYRS for Methyl-O-tyrosine is 100 fold higher than that for tyrosine. The physiological concentration of tyrosine in the *E. coli* cytoplasm is estimated to be 80 μM, which is far below the K_m i.e 5833 μM of the mutant TYRS for tyrosine. In case of Methyl-O-tyrosine the concentration of the methyl tyrosine in treated cells is comparable or greater than the K_m which is 443 μM (Wang et al., 2006).

3. Misacylation of the unnatural amino acid with a native amino acyl tRNA synthetase to the tRNA.

Cells contain a distinct amino acyl tRNA synthetase for each of the 20 amino acids. There is a possibility of the catalysis of the acylation by a native aminoacyl tRNA synthetase rather than the specific orthogonal synthetases (Furter, 1998). The Schultz group claim that this has been taken care of by the screening supported by the structure based specificity assays. However in case this happens, the labeled protein will still be formed.

4. Expression levels

It has been observed that the level of expression when compared to the wild type protein is less. This is a result of cumulative factors including ribosomal pause at the stop codon, the intracellular availability and proximity of the tRNA, synthetase and the unnatural amino acid. The catalysis or the acylation efficiency, cross reactivity, mis acylation and miscarriage with the endogenous entities at each step are some of the aspects that will influence the efficiency of the incorporation.

Over expressing the protein of interest by placing it under the control of an over expressing induction system like T7 promoter will lead to the formation of higher amounts of the truncated products due to the imbalance between the expression level and the ribosomal pause. Furter has observed two third expression of the protein when compared to the wild type expression level (Furter, 1998).

5. Readthrough

Translation usually stops when the stop codon enters the A site in the ribosome. *E. coli* has two release factors (RF) that facilitate in the hydrolysis of the polypeptide chain form the last tRNA. However the stop efficiency is not 100 % and readthrough does occur. Readthrough has been observed in *E. coli* in response to the amber stop codon. $tRNA_{Gln}$

has been found to read the stop codon to incorporate glutamine in its place (Nilsson et al., 2003). *E. coli* has two glutamine tRNAs, one decoding for CUG and the other for CAG. This tRNA can base pair with UAG is a G-U pair is formed between the first nucleotide in the codon and the third in the anticodon.

In *E. coli* the release factor (RF1) acts on UAG stop codon. The magnitude of suppression of the stop codon has been found to be dependent on the position of the codon based on the capacity of wobble pairing of the third position in the codon (Miller, 1991; Martin et al., 1988).

2.8 Examples of labeling proteins using the fluorescent amino acid

7-hydroxycoumarinyl-ethylglycine has been used for labeling proteins for further studies:

1. Myoglobin (PDB 105M) with 7-hydroxycoumarinyl-ethylglycine (Wang et al., 2006).
2. The serine 4 position was labeled with the unnatural amino acid. 'UCU' for serine is mutated to UAG. Not conserved and not involved in functional aspects.
3. Schultz group have applied this amino acid and labeled antibody 5C8 and eventually tested antigen antibody interactions which supports the workability of this technique and this particular amino acid (Mills et al., 2009).
4. Aspartate transcarbamoylase (1NBE), a 153 residue protein with 7-hydroxycoumarinyl-ethylglycine in a position structured in a surface exposed loop at glycine 52 by Mendes et al. (2010).

2.9 Chemical synthesis of L-(7-hydroxycoumarin-4-yl) ethylglycine

The fluorescent amino acid is commercially unavailable and was synthesized in house as follows.

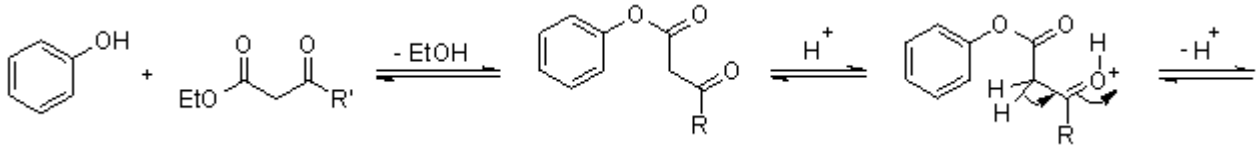
2.9.1 Mechanism & Synthesis:

The amino acid is synthesized following the Pechman condensation method that allows the synthesis of coumarins by reaction of phenols with β-keto esters.

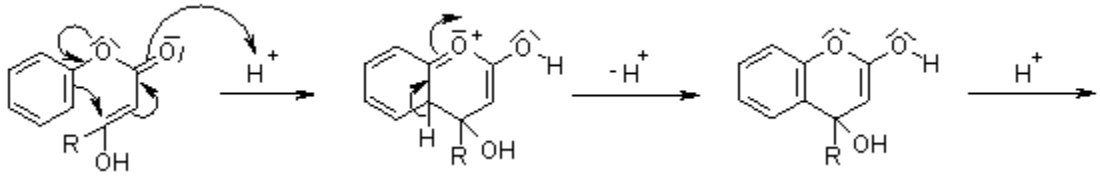
The general mechanism of the reaction is given below.

- **Mechanism**

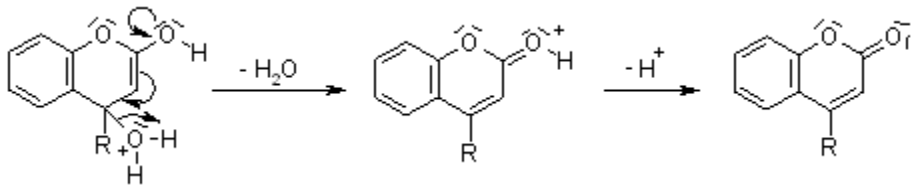
The reaction is conducted with a strong Brønstedt acid such as methanesulfonic acid or a Lewis acid such as $AlCl_3$. The acid catalyses transesterification as well as keto-enol tautomerisation:



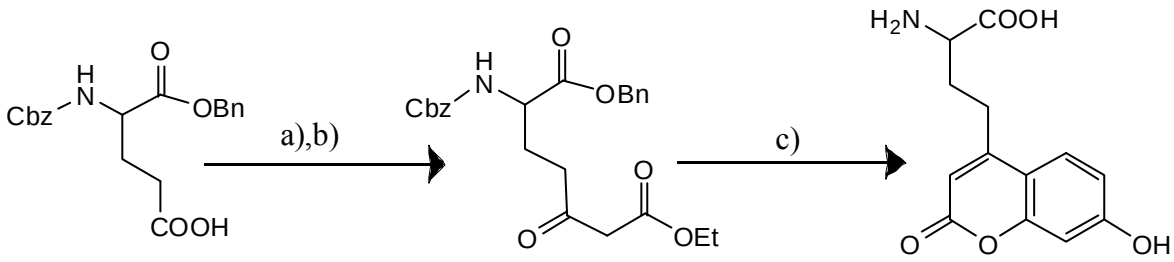
A Michael Addition leads to the formation of the coumarin skeleton. This addition is followed by rearomatisation:



Subsequent acid-induced elimination of water gives the product:



Following is the general scheme for the synthesis of L-(7-hydroxycoumarin-4-yl) ethylglycine (Brun et al., 2004)



N-α-Cbz-L-glutamic acid α-benzyl ester

(2S)-2-benzyloxycarbonylamino -5-oxo-heptanedioic acid 1-benzyl ester 7-ethyl ester

L-(7-hydroxycoumarin-4-yl)ethylglycine

Where, a) N,N'-Carbonyldiimidazole, RT, 2 Hrs ;b) Ethyl Magnesium malonate#, RT, overnight;c) Resorcinol, methanesulfonic acid, RT, 2 Hrs

Ethyl magnesium malonate is synthesized from ethyl potassium malonate and magnesium chloride and isopropanol.

The coumarinyl amino acid was synthesized by a two step process first converting N-α-Cbz-L-glutamic acid α-benzyl ester into the side chain β-keto ester which was then reacted with resorcinol in methanesulfonic acid to afford the amino acid product. The synthesis was carried out as below.

2.9.2 Synthesis methods

I. Synthesis of Ethyl magnesium malonate

1. 1.5 M Magnesium chloride + 3 M Ethyl potassium malonate + 5 volumes Isopropanol
2. Filtration and Ethyl magnesium malonate (product) dried in vacuo

II.Synthesis of (2S)-2-benzyloxycarbonylamino-5-oxo-heptanedioic acid 1-benzyl ester 7-ethyl ester

1. Z-Glu-Obzl (5 g,13.46 mmol) dissolved in 50 ml dry THF at RT (4 ml/mmol THF)
2. Carbonyl diimidazole (1.1 eq) added slowly,
3. Mixture stirred for 2 Hrs at RT, cooled to 0 ^{0}C
4. Ethyl magnesium malonate (0.55 eq) added & stirred at room temperature overnight
5. Product extracted with ether (15-20 ml/mmol of amino acid)
6. Washed with 10 % sodium bicarbonate, water and brine
7. Residue purified by flash chromatography on silica gel (ethyl acetate:hexanes 1:1)
8. Concentrated on rotary evaporator → white solid (40 % yield).
9. 1H NMR spectra recorded on Bruker AMX-400 with chemical shifts reported relative to tetramethylsilane.
10. LC-MS (ESI)

III.Synthesis of L-(7-hydroxycoumarin-4-yl) ethylglycine

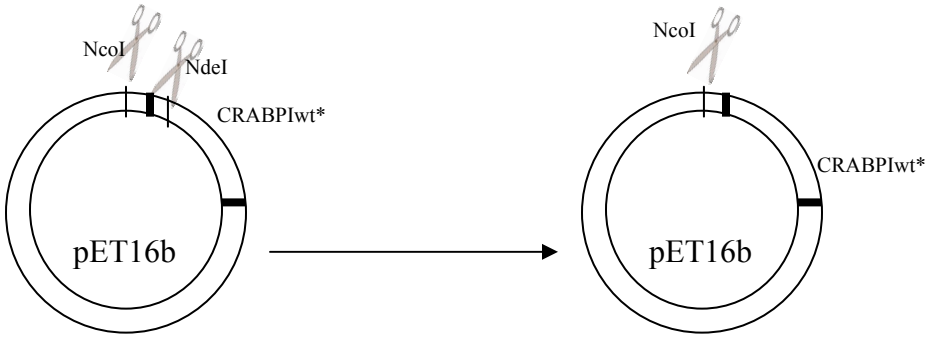
1. (2S)-2-benzyloxycarbonylamino-5-oxo-heptanedioic acid 1-benzyl ester 7-ethyl ester (2.35 g,5 mmol) added slowly to resorcinol (3 g, 25 mmol) in methanesulfonic acid (20 ml)
2. Stirred 1-2 Hr at room temperature
3. Deep-red homogeneous mixture taken into 5 volumes cold ether(-30 ^{0}C)
4. Centrifuged for 20 minutes at 3600 × g
5. Discard ether
6. Precipitate washed with cold ether
7. Centrifuged 20 minutes at 3600 × g
8. Dissolved in water, filtered and Lyophilised
9. Product purified by reverse phase HPLC → brown solid in 50 % yield (YMCAA12S052503WT column, 12 ml/min flow rate, from 10 % to 90 % CH_3CN, 0.1 % TFA (w/v) in water, over 12 min) and LC-MS (ESI).

2.10 Molecular biology techniques

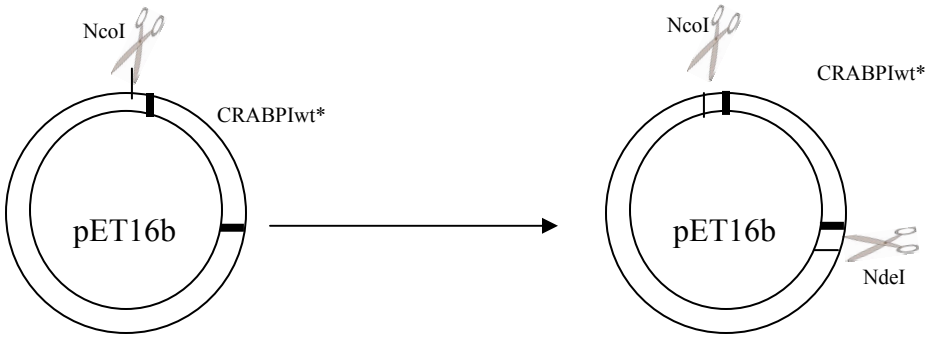
2.10.1 Cloning

The pBAD/JYAMB-4TAG-Myo and pBK-CouRS-D8 plasmids were sent by the Schultz group (Wang et al., 2006). The plasmids were transformed in DH5α cells and tested by sequencing and agarose gel run.

To clone the CRABP I gene into pBAD/JYAMB-4TAG-Myo in place of myoglobin the NcoI and NdeI restriction sites were exploited. The CRABP I wt* gene in the pET16bCRABPIwt* plasmid has the NcoI and NdeI restriction sites before the and after the start site for the CRABP I wt* gene while the myoglobin gene in the pBAD/JYAMB-4TAG-Myo vector has the NcoI and NdeI flanking sites. The NdeI site hence was removed from its original place and reintroduced after the CRABP I gene by a two step site directed mutagenesis. The first step was the removal of the NdeI site as depicted in the schematic below.

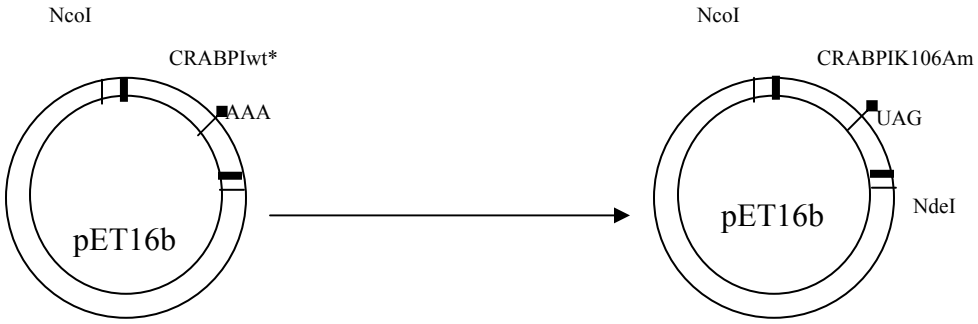


The second step was to introduce the NdeI site after the CRABP I wt* as shown below.



2.10.2 Site directed mutagenesis of K106 to amber codon:

The primers for site directed mutagenesis for Lysine 106 to amber codon (AAA to UAG) in CRABPIwt* (pET16bCRABPIwt* plasmid) were designed and obtained. Site directed mutagenesis by PCR was hence performed. The subsequent product was transformed into DH5α cells after which the plasmid was extracted and sent for sequencing with the T7 forward primer. The schematic below represents the site directed mutagenesis step performed.

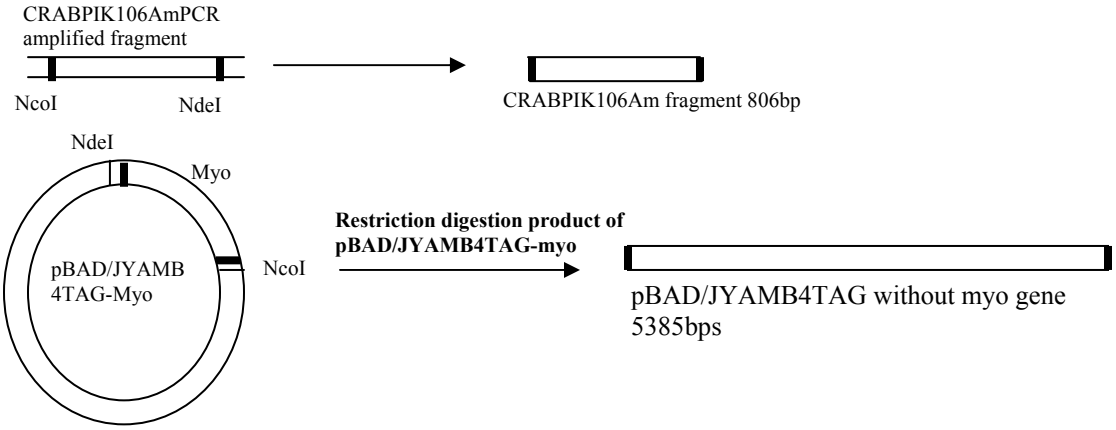


2.10.3 Cloning of the CRABPIK106Am gene into pBAD/JYAMB-4TAG-Myo plasmid:

The pETCRABPIK106Am plasmid was subjected to PCR amplification with the T7 forward and reverse primers. The product of which was a 944 bps in length fragment. The PCR product is run on an agarose gel to check the presence of the amplified product. The following procedures were followed to obtain the desired product.

2.10.4 Restriction digestions:

The pBAD/JYAMB-4TAG-Myo and the PCR amplification product of the CRABPIK106Am gene were digested with the NcoI and the NdeI restriction enzymes to obtain similar sticky ends. The whole product was loaded onto an agarose gel and the required fragments were excised from the gel and purified by the Qiagen gel extraction procedure and eluted in warm water. The following schematics depict the products obtained as a result of the restriction digestions of the two fragments.



2.10.5 Ligation of the restriction digestion fragments:

The CRABPIK106Am fragment of size 806 bp was ligated to the pBAD/JYAMBTAG frame of 5385 bps. The ligation was performed with a ration of 1:3 (vector:insert) for 1Hr at room temperature. The product was transformed into DH5α cells and the eluted product of plasmid extraction was sent for sequencing to confirm ligation and to check whether the desired product is formed. The plasmid obtained was hence named as pBAD-CRABPIK106Am. The size of the plasmid is 6415 bps. Similar method was followed to obtain the pBAD-CRABPIwt* plasmid without the K106Am mutation.

2. 11 Preliminary studies prior to *in vivo* labeling of CRABP I

2.11.1 Permeability of the unnatural amino acid

The permeability of the 7-hydroxycoumarinyl 4-ethyl glycine was studied by a fluorescence assay. A single colony of BL21 (DE3) cells harboring the pBAD-CRABPIK106Am and the pBK-CouRS-D8 plasmids was grown in a 5ml starter culture with the antibiotics tetracycline and kanamycin at their respective concentrations. The culture was allowed to grow to its confluence overnight at 37^0C. This culture was used to inoculate a 50 ml 2XYT medium to a starting optical density at 600 nm of 0.09. The cells were allowed to grow at 37 ^{0}C at 220 rpm till the optical density at 600 nm reaches 0.68, which took four and a half hours. At this time the culture was induced with 0.2 % arabinose and the 7-hydroxycoumarinyl 4-ethyl glycine at a concentration of 1 mM was

also added (Wang et al., 2006). The cultures were induced for 4 Hrs and then harvested. Cultures at intervals of 2 Hrs were taken to monitor the fluorescence. The culture samples were washed in phosphate buffer thrice to remove the external fluorescent molecules and then centrifuged at 3500 rpm for 10 minutes at 4 ^{0}C and washed thrice with autoclaved deionized water to open up the cells and check the fluorescence of the cytoplasm. The pellet was resuspended in deionized water and the fluorescence was measured in a spectrofluorometer. The samples were excited at 366 nm and the emission from 350 nm to 500 nm was collected.

2.11.2 Toxicity of the unnatural amino acid

Effect of 1 mM 7-hydroxycoumarinyl 4-ethyl glycine synthesized in lab on the growth pattern of *E. coli* cells was assessed. This assay provided us with information regarding the differences in growth patterns when compared to growth in the absence of the amino acid. The cell viability was measured by optical density measurement at 600 nm and also by plating the culture at 1 Hr time intervals. The growth curves in comparison with the wild type were plotted.

2.11.3 Optimization of expression conditions

Labeling of CRABP I was conducted using a final 100 ml volume culture. The plasmids pBAD-CRABPIK106Am and pBK-CouRS-D8 were co-transformed into BL21 (DE3) cells. A single colony was picked for preparation of the overnight starter culture in 5 ml of 2XYT medium with the tetracycline and kanamycin antibiotics. This was used to inoculate 100 ml of the 2XYT medium supplemented with the antibiotics and 1 mM of the fluorescent amino acid. The cells were allowed to grow till an O.D 600 nm of 0.5 at which point they were induced with 0.2 % arabinose according to the protocol (Wang et al., 2006). The cells were harvested after 12 Hrs. BPER II reagent was used to lyse the cells and separate the soluble and insoluble fractions. The fluorescent and coomassie stained gels were observed.

The wild type CRABP I gene cloned into the pBAD vector was used for this purpose. The pBADCRABPIwt* was transformed into BL21 (DE3) cells and used for these

assays. The two major parameters are -Time of induction and Concentration of inducer-arabinose.

To test the cells were grown as above and at mid-log phase the cells were induced with 0.002 %, 0.02 % and 0.2 % arabinose. Samples were taken at hourly intervals for 6 Hrs and the soluble fraction was analysed for the presence of the protein. It is observed that the expression of the CRABP I wt* protein was highest at 0.2 % arabinose induction.

2.12 Expression of labeled protein

To test if suppression of the amber codon protein takes place without the addition of the unnatural amino in the medium and to also assay and optimize the level of expression of the labeled protein the nonsense suppression system for Methyl-O-tyrosine was procured from the Schultz lab. The procedure followed for the expression of the protein was as follows. 10 and 50 ml cultures were used for expression of CRABPIK106Am. The pET16bCRABPIK106Am plasmid and the pSupJYRS plasmids were co-transformed into BL21 (DE3) cells and single colonies of cells were inoculated into 2XYT medium with ampicillin and chloramphenicol for the preparation of the overnight starter culture. This was used to inoculate the growth culture to start with an OD at 600 nm of 0.2. It took 3 hrs for the culture to attain an OD of 0.7 at which point the culture was induced with 400 uM IPTG and 1 mM methyl-O-tyrosine was added. The cells were allowed to grow at 37 ^{0}C, 220 rpm for 0.5 Hr. After which the culture was split to two 5 or 25 ml cultures. Both were grown at 30 ^{0}C. One was harvested after 4 Hrs and the other after overnight induction.

Plasmid stability test was performed on LB, LB+ampicillin+chloramphenicol+IPTG, LB+ampicillin+chloramphenicol, LB+IPTG agar plates in case of overnight induction to confirm the presence of both the plasmids during the expression of the protein. It was confirmed that the plasmids were stable during the whole period.

The harvested cultures were lysed with BPER II reagent and the lysates were analyzed by coomassie staining, western blotting using anti-CRABP I antibodies and anti-His tag antibodies.

CHAPTER 3

MOLECULAR BIOLOGY

3.1 Results

3.1. A Cloning

3.1. A. 1 Removal of the NdeI restriction site from pET16bCRABPIwt* plasmid

The NdeI recognition site present after the start of the CRABP I wt* gene in pET16bCRABPIwt* plasmid was removed by site directed mutagenesis as mentioned in chapter 2. The resultant product is shown in Fig 3.1.a. The resultant plasmid was sent for sequencing and the result analyzed for the absence of the NdeI site at the previous location. It was observed that the NdeI site was successfully deleted from the earlier site.

3.1. A. 2 Introduction of the NdeI restriction site in pET16bCRABPIwt* plasmid

The NdeI recognition site was introduced after the CRABP I wt* gene by site directed mutagenesis as mention in the previous chapter. The product obtained was run on an agarose gel as shown in Fig 3.1.b and subsequently sent for sequencing, the result of which was analyzed and proven to have the NdeI site after the gene at the desired location.

3.1. A. 3 Site directed mutagenesis of K106 to amber codon

The Lysine 106 position in CRABP I wt* (pET16bCRABPIwt* plasmid) was successfully mutated from AAA to UAG by Site directed mutagenesis by PCR as mentioned in chapter 2. The plasmid obtained as a result was subjected to sequencing and checked for the mutation at the right position. The mutagenesis hence worked and the plasmid, pETCRABPIK106Am, obtained was further subjected to the following steps.

3.1. A. 4 Cloning of the CRABPIK106Am gene into pBAD/JYAMB-4TAG-Myo plasmid

The pETCRABPIK106Am plasmid was PCR amplified with the T7 forward and reverse primers and the product obtained was 944 bps in length (Fig 3.1.c).

The pBAD/JYAMB-4TAG-Myo and the PCR amplification product of the CRABPIK106Am gene were digested with the NcoI and the NdeI restriction enzymes to obtain similar sticky ends. The restriction digestion products of the former are observed in Fig 3.1.d.

3.1. A. 5 Ligation of the restriction digestion fragments

The CRABPIK106Am fragment of size 806 bps was ligated to the pBAD frame of 5385 bps as per the procedure mentioned in chapter 2.

The plasmid thus obtained was analyzed by sequencing and was named pBAD-CRABPIK106Am. The size of the plasmid is 6415 bps.

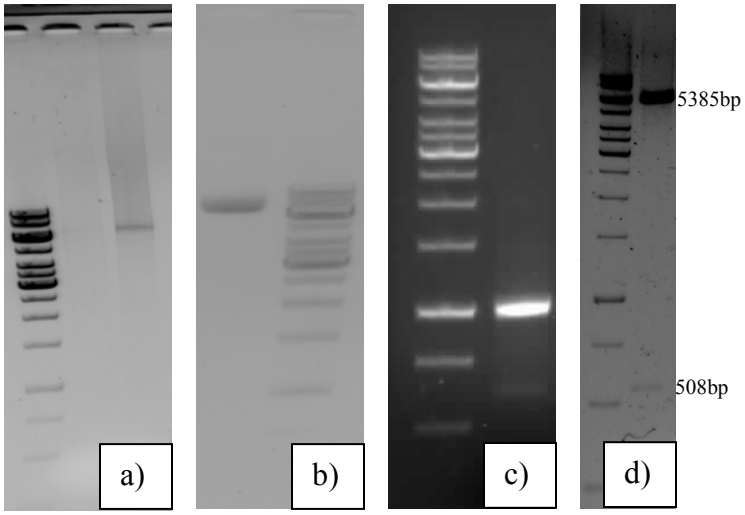


Fig 3.1 Agarose gels representing each stage of the plasmid preparation process. a)Removal of the NdeI restriction site from pET16bCRABPIwt* plasmid b) Introduction of the NdeI restriction site from pET16bCRABPIwt* plasmid c) 944 bps long PCR product of CRABPIK106Am gene d) Restiction digestion fragments of pBAD/JYAMB-4TAG-Myo, 53835 bps and 508 bps respectively.

CHAPTER 4

SYNTHESIS OF THE FLUORESCENT AMINO ACID

4.1 Results

4.1. A Chemical synthesis of L-(7-hydroxycoumarin-4-yl) ethylglycine

The synthesis procedure was followed as mentioned in chapter 2. The intermediate product was subsequently obtained and tested by LC-MS (ESI) and 1H-NMR (400 MHz,CDCl3). Expected mass: 465.0 Da (M+Na+), Observed: 464.1 Da (M+Na+). The obtained NMR spectrum was as expected (Wang et al., 2006, Brun et al., 2004).The figures of the LC-MS and 1H-NMR are provided in Fig. 4.1 and 4.2.

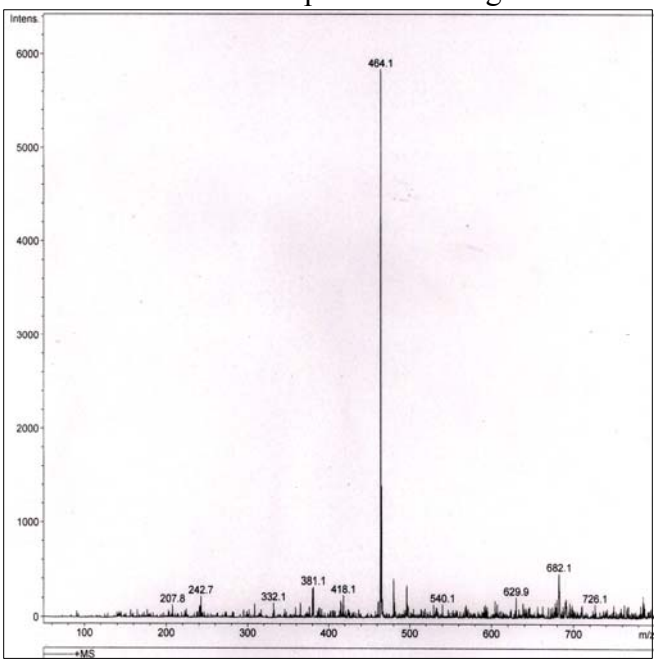


Fig 4.1 LC-MS (ESI) spectrum of the intermediate product ((2S)-2-benzyloxy carbonyl amino-5-oxo-heptanedioic acid 1-benzyl ester 7-ethyl ester) in the synthesis of the final fluorescent amino acid. Expected mass:465.0 Da, found:464.1 Da

The NMR spectrum of the intermediate compound was also procured. The expected peaks were compared with the obtained peaks (Fig 4.2). The intermediate compound was hence obtained.

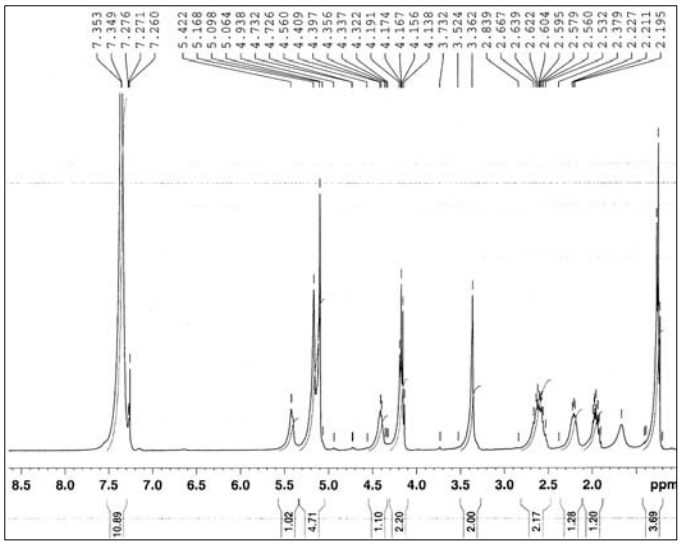


Fig 4.2 1D H^+NMR spectrum of the intermediate compound (2S)-2-benzyloxycarbonyl amino-5-oxo-heptanedioic acid 1-benzyl ester 7-ethyl ester

The intermediate product was further subjected to the next step to obtain the fluorescent amino acid. The product was purified using a preparative C-18 column with a 5-45 % CH_3CN gradient as represented in Fig 4.3. The fractions were collected and that from the 11^{th} to 14^{th} minute was rotavaporized and sent for LC-MS.

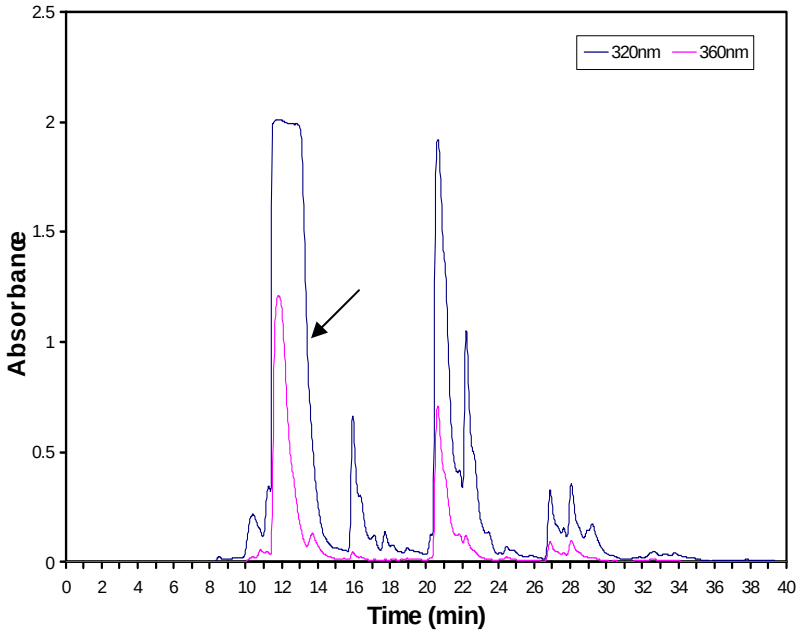


Fig 4.3 HPLC purification of the fluorescent amino acid. The spectrum is shown in this figure where the fractions pointed were collected.

ESI-MS crude product has an expected mass of 264.08 Da. The observed mass was 264.0 Da as represented in the ms spectrum in Fig 4.4.

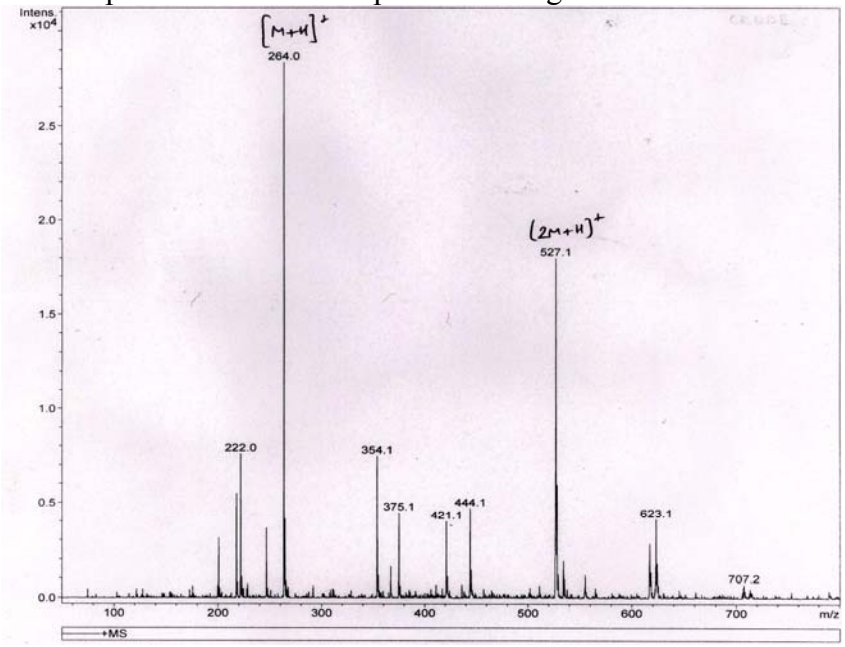


Fig 4.4 LC-MS (ESI) spectrum of the crude fluorescent amino acid prior to HPLC purification. Expected mass:264.08Da, found:264Da

The HPLC purified fractions were then also tested by LC-MS (ESI) to obtain a purified compound of required molecular mass as shown in Fig 4.5.

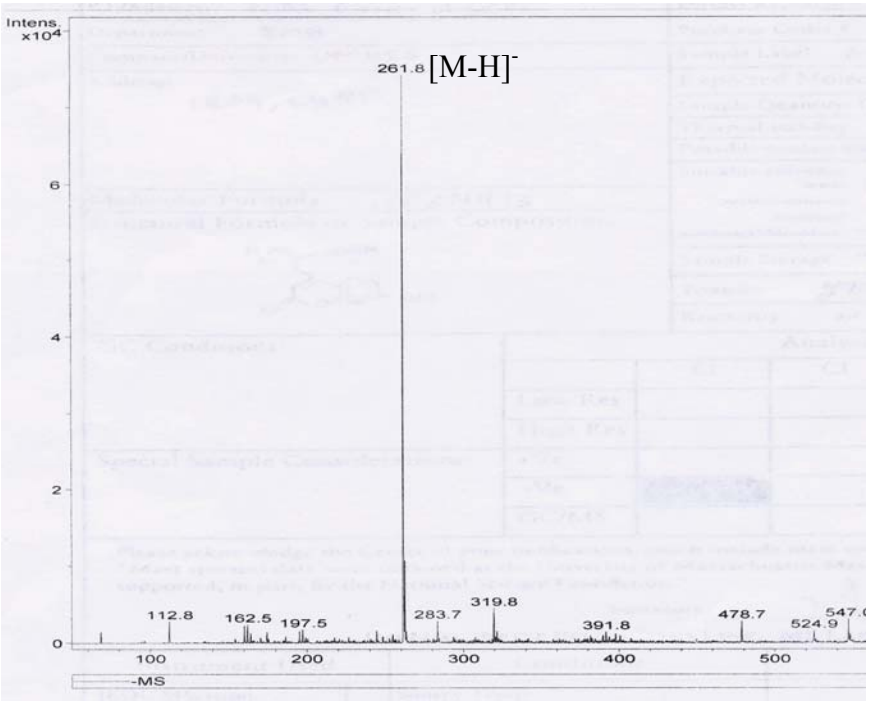


Fig 4.5 LC-MS (ESI) spectrum of the final fluorescent amino acid product. Expected mass: 264.08 Da, found:261.8 Da

An absorbance spectrum of the product was performed in 100 mM sodium phosphate buffer at pH 7.4. The spectrum obtained is shown in Fig 4.6.

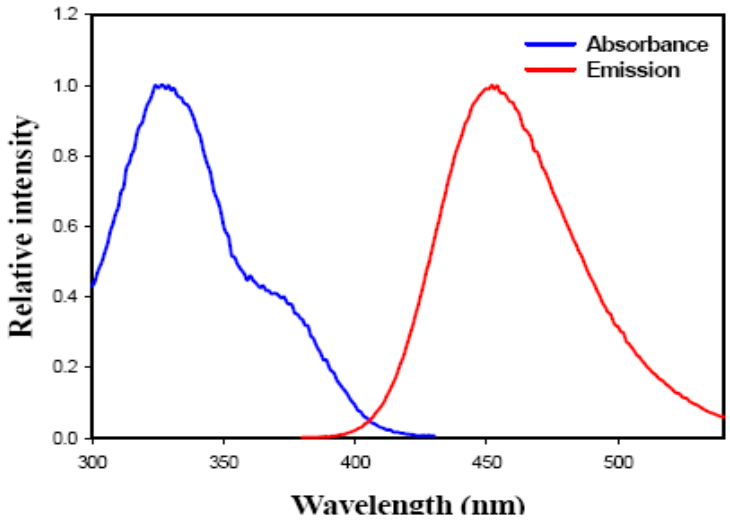


Fig 4.6 Absorbance spectrum of the purified fluorescent amino acid in phosphate buffer pH7.4. Excitation maximum at 340 nm and emission maximum at 450 nm.

The fluorescent amino acid was hence synthesized. The yield of the amino acid was 500 mg from 2 gm of starting compound.

4.1. B Tests to confirm the presence of free amino acid in the product

Ninhydrin test was performed on a TLC plate. The purity of the sample could also be tested with this. Butanol:Acetic acid:Ethylacetate:water (1:1:1:1) solvent mixture was determined best for separation of the compound. The amino acid sample was dissolved in it and loaded on the TLC plate which was then placed in a saturated chromatography chamber with the solvent. The ninhydrin reagent was prepared in n-Butanol and sprayed on the air dried TLC plate. The TLC plate was visualized using a hand UV torch. A picture of the plate was also taken (Fig 4.7 a). A clear band is observed in the case of the fluorescent amino acid and the band turned positive for ninhydrin.

4.1. C To test the overall charge of the amino acid:

Pure 7-hydroxycoumarin or umbelliferone is weakly acidic, soluble in hot water and acetic acid and has a molecular weight of 162.14 Da. The nature of the synthesized amino acid was tested using agarose gel. The fluorescent amino acid was loaded with glycerol onto an agarose gel (1 %) with wells made at the center of the gel. Current was applied and the trace of the fluorescent molecule was followed using the hand UV torch. The amino acid was found to be moving towards the positive electrode suggesting the anionic nature of the amino acid (Fig 4.7 b).

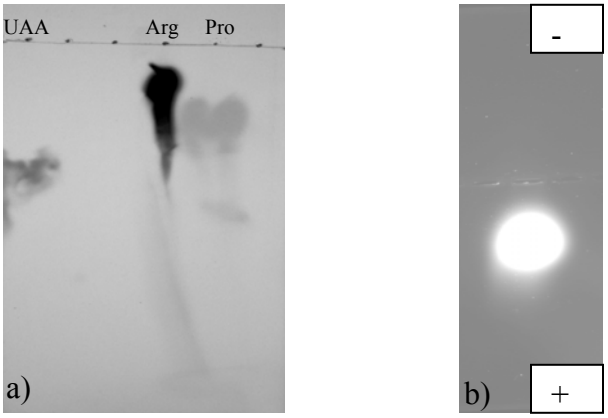


Fig 4.7 Tests to check the free amino acid and overall charge of the amino acid a) ninhydrin assay to test the presence of the free amino acid. UAA-Unnatural amino acid;Arg-Arginine; Pro-proline, the product hence is positive for ninhydrin; b) the electrophoretic mobility of the free amino acid in an agarose gel. The free fluorescent amino acid is observed to be negative in charge.

This was further confirmed by fluorometric assays of the compound in different solvents and pH conditions. Scientists (Colambano and Troccoli, 1985) postulate that 7-hydroxycoumarin exists in equilibrium among cationic, anionic and neutral species. Each of these species can be differentiated by the emission and excitation maxima which are specific. The neutral species emits maximum at 400 nm when excited at 325 nm. The

cationic form emits at 431 nm when excited at 345 nm. The anionic form emits at 453 nm when excited at 366 nm.

The compound was dissolved in water and the emission spectra were collected for three different excitation maxima (Fig 4.8 a).

It has been found from the emission spectra that the compound is predominantly in the anionic form. pH can also have an effect on the fluorescence emission of the compound is shown. The intensity of fluorescence decreases with increasing pH (Fig 4.8 b) (Colombano and Troccoli, 1985), thus making it a good candidate for being a pH sensor (Budisa et al., 2002).

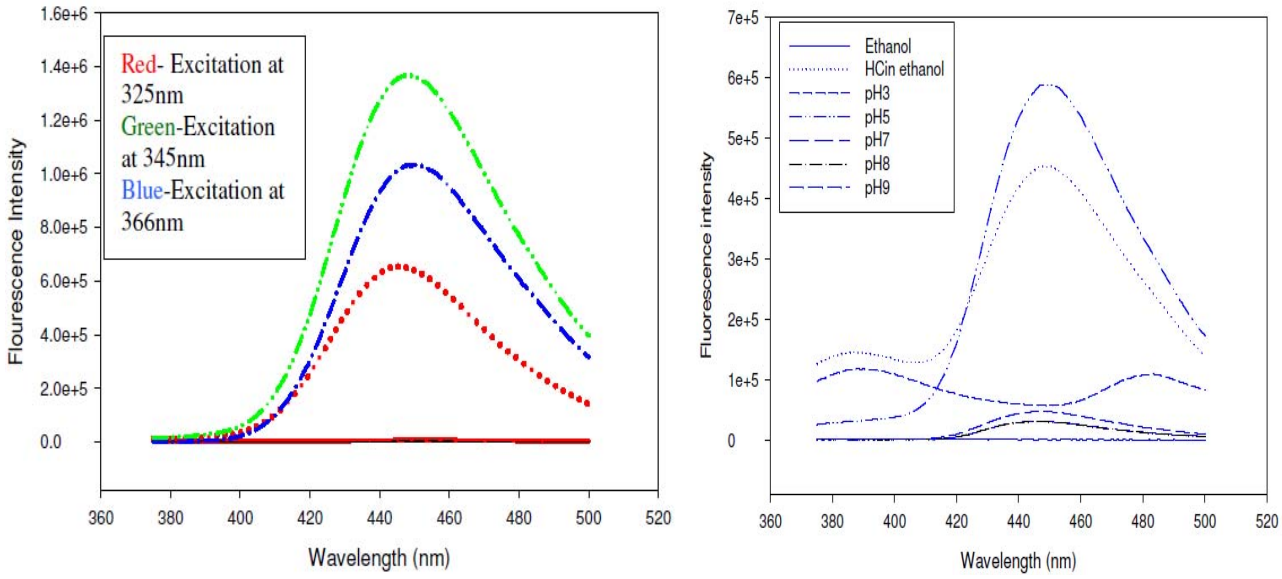


Fig 4.8 Emission scans at various excitation wavelengths and pH. a)Emission scans of the fluorescent amino acid at different excitation wavelengths of 325, 345 and 366 nm. b) Emission spectra of the fluorescent amino acid at different pH conditions while excited at 366 nm.

This is because of intramolecular charge migration and can be used as a tool in protein folding, protein membrane interaction, protein-protein interaction, protein-ligand interactions studies and other applications (Mills et al., 2009; Gök et al., 2008;

Uttamapinant et al., 2010). Hence, the labeled proteins can act as protein based molecular sensors.

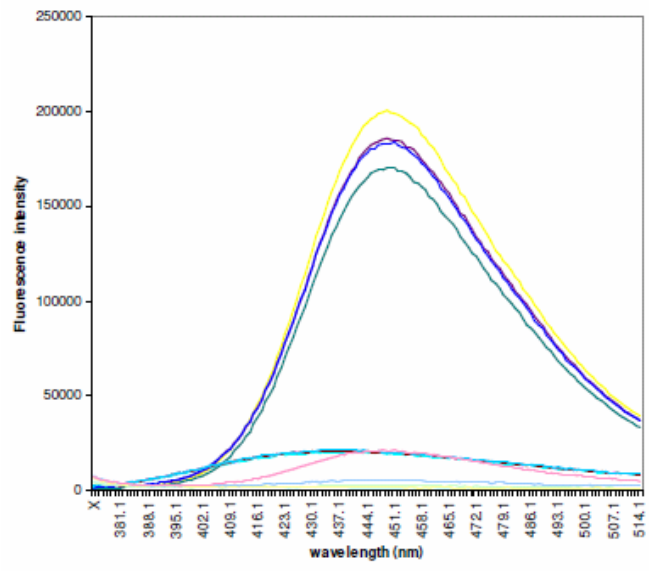
CHAPTER 5

PRELIMINARY STUDIES PRIOR TO LABELING

5.1 Results

5.1. A Permeability of the unnatural amino acid into the *E. coli* cells

The emission maximum for the 2XYT medium alone is 400 nm. With the 7-hydroxycoumarinyl 4-ethyl glycine the emission maximum shifts to 450 nm. The pellet has no fluorescence before addition of the 7-hydroxycoumarinyl 4-ethyl glycine and even after 2 Hrs of 7-hydroxycoumarinyl 4-ethyl glycine addition. In the culture with the 7-hydroxycoumarinyl 4-ethyl glycine added at the time of induction, the pellet appears to be non-fluorescent even 2 Hrs later. The fluorescence is observed after 4 Hrs depicting that increased induction time is effective for permeability of the 7-hydroxycoumarinyl 4-ethyl glycine into the cells (Fig 5.1). Hence, it can be concluded that the permeability of the fluorescent amino acid increases with time and an induction period greater than 4 Hrs is best suitable for labeling to be efficient if addition of the fluorescent amino acid is done at the time of induction. There is evidence that the growth of the bacteria is also affected after the addition of the fluorescent amino acid at the time of induction as compared to at the start of the culture.



Yellow- 4Hrs post induction and addition of 7-hydroxycoumarinyl 4-ethyl glycine
Purple-Culture at induction time (with 7-hydroxycoumarinyl 4-ethyl glycine added at the start)
Blue-4Hrs post induction of the culture (with 7-hydroxycoumarinyl 4-ethyl glycine added at the start)
Green-2XYT+7-hydroxycoumarinyl 4-ethyl glycine
Cyan-2XYT
Red-Culture with no 7-hydroxycoumarinyl 4-ethyl glycine
Pink-Culture at start (with 7-hydroxycoumarinyl 4-ethyl glycine added at the start)
Light blue- Ethanol

Fig 5.1 Permeability of the fluorescent amino acid, 7-hydroxycoumarinyl 4-ethyl glycine into BL21 (DE3) cells. The emission spectra are depicted by different colors which are mentioned in the legend on the right.

5.1.B Toxicity of 7-hydroxycoumarinyl 4-ethyl glycine

Effect of 1mM 7-hydroxycoumarinyl 4-ethyl glycine synthesized in lab on the growth pattern of *E. coli* cells was assessed. This assay provided us with information regarding the differences in growth patterns when compared to growth in the absence of the amino acid. The cell viability was measured by optical density measurement at 600 nm and also by plating the culture at 1 Hr time intervals. The growth curves when compared to the wild type were plotted (Fig 5.2).

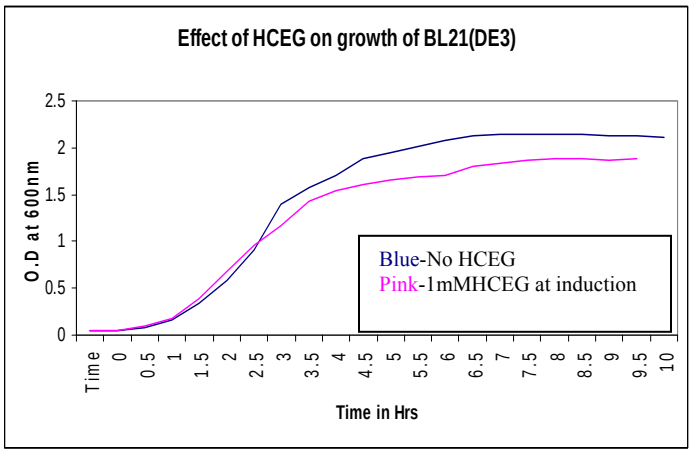


Fig 5.2 Effect of the fluorescent amino acid on the growth of BL21(DE3) cells. The O.D at 600 nm was measured at 30 minutes interval. 1mM of the fluorescent amino acid was added at induction time of 3 Hrs

It is observed that there is a difference in the growth pattern after induction. The number of colonies that grew on LB agar after the induction period and addition of the fluorescent amino acid when compared to the culture without the addition of the fluorescent amino acid also had a similar effect. These studies show that the fluorescent amino acid has an influence on the growth of the bacteria. There is a slight retardation of the growth after the addition of the fluorescent amino acid. This has to be taken into consideration in terms of regulation of the concentration of the amino acid provided to the culture.

5.1. C Preliminary expression of labeled protein

Preliminary expression studies were carried out using the method mentioned in chapter 2 and the supernatant samples were loaded and run on a SDS PAGE gel which was coommassie stained. Prior to staining, the gel was observed under UV light and faint bands were observed, however an image was not recorded. The coommassie stained gel is shown in Fig 5.3. From the gel it is clear that a corresponding band is observed for the CRABPK106Am protein when expressed with the fluorescent amino acid as pointed with the arrow. However, there was no band in the previous lane, which was a repeat of the same expression test (duplicate). The last lane to the right shows the wild type CRABP I expressed under standard conditions. Hence, the result seemed very inconsistent.

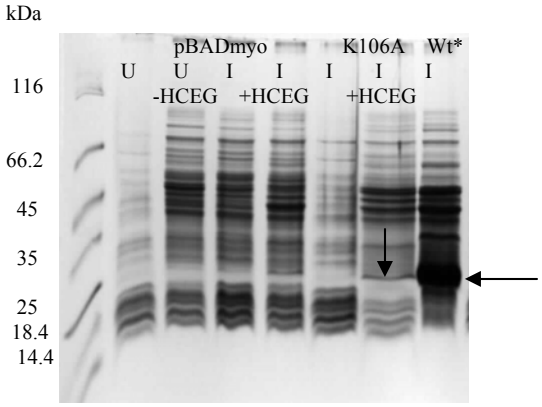


Fig 5.3 Expression test of CRABP I labeled with the fluorescent amino acid at the K106 position; Coomassie stained SDS-PAGE gel;I=Induced, U=uninduced. Evidence of labeled protein has to further be observed clearly due to lack of consistency and optimization of expression.

5.1.D Optimization of the conditions for protein expression

Since the expression level of the protein seemed to very low the parameters for expression of protein were altered and assayed. The wild type CRABP I gene cloned into the pBAD vector was used for this purpose. The pBADCRABPIwt* was transformed into BL21 (DE3) cells and used for these assays. The two major parameters are -Time of induction and Concentration of inducer-arabinose.

To test these cells were grown as above and at mid-log phase the cells were induced with 0.002 %, 0.02 % and 0.2 % arabinose. Samples were taken at hourly intervals for 6Hrs and the soluble fraction was analysed for the presence of the protein. It is observed that the expression of the CRABP I wt* protein was highest at 0.2 % arabinose induction. However, the protein expression level was not high as seen in the Fig 5.4. The increase in the level of the protein can be observed with time i.e 4 Hrs.

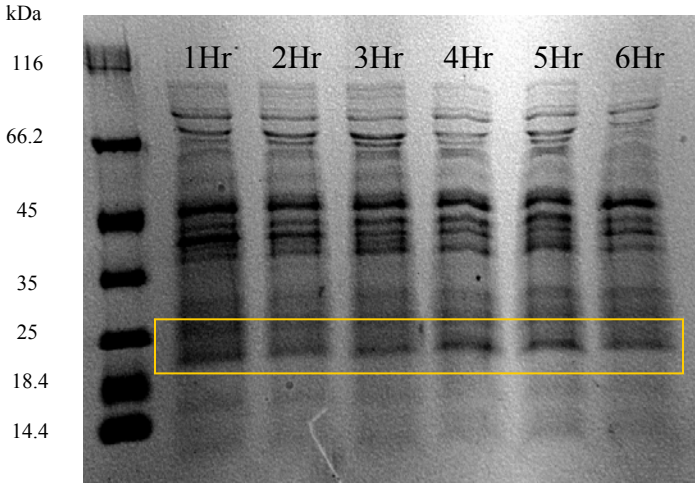


Fig 5.4 Optimization of the arabinose concentration for induction. This is a Coomassie stained SDS-PAGE gel of CRABP I wt* expressed in the pBAD system, with hourly samples after induction with 0.2 % arabinose. An increased expression is observed over time as marked within the rectangle.

5.1. E Use of an alternative suppression system

On performing a small scale labeling experiment using the same system mentioned above it was observed that no labeled protein was obtained. The system hence did not have consistency over different culture volumes. The question of optimizing the conditions for suppression to occur arose. To test and study if suppression was taking place a different system was to be used keeping in mind the low amount of the fluorescent amino acid present. The Methyl-O-tyrosine system seemed to be suitable for studying the

suppression. The plasmid required for this system to work was obtained from the Schultz group. The methyl-O-tyrosine is easier for studies when compared to the fluorescent amino acid as it is commercially available.

5.1.F Toxicity and effect of Methyl-O-Tyrosine on growth of bacteria

Similar growth effect assay, as done for the fluorescent amino acid was conducted for the Methyl-O-tyrosine. 2 mM of methyl-O-Tyrosine was added at the time of induction in one set of culture. The growth curves suggest that the growth of the bacteria is effected by the concentration of the amino acid in the medium. The amino acid was added in both the cases at the time of induction i.e. at 3.5 Hrs. It is observed that the growth of the bacteria in 1 mM concentration of the amino acid is similar to the control and hence has no detrimental effect on the cells (Fig 5.5).

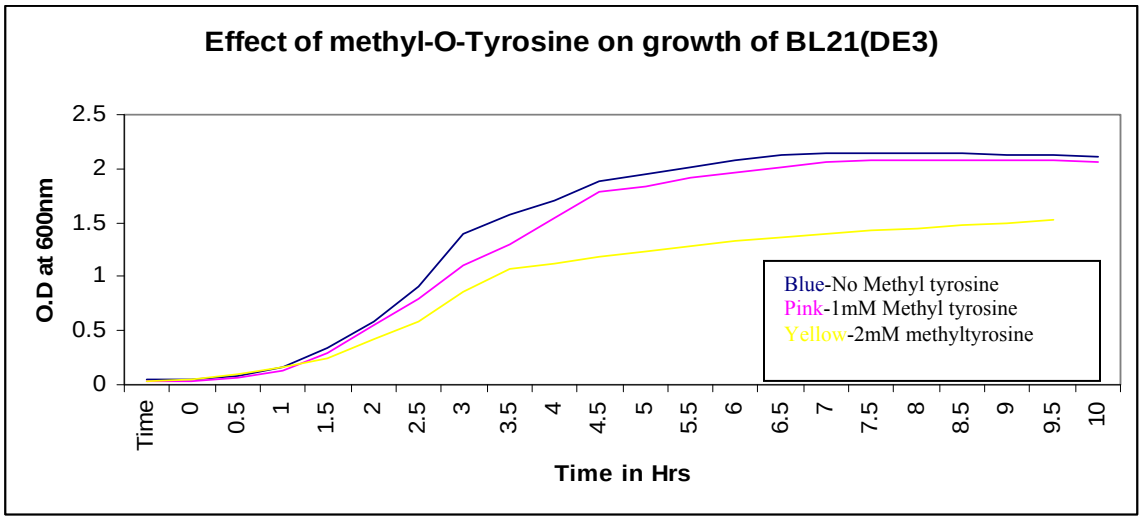


Fig 5.5 Effect of growth of BL21(DE3) cells by methyl-O-tyrosine by growth curve experiment. Higher concentrations of the amino acid is harmful to the cells, but a concentration of 1 mM is optimum for growth.

CHAPTER 6

EXPRESSION OF THE LABELED PROTEIN

6.1 Results

6.1. A Suppression assay-Expression of protein labeled with Methyl-O-tyrosine

To test if suppression of the amber codon protein takes place without the addition of the unnatural amino in the medium and to also assay and optimize the level of expression of the labeled protein the nonsense suppression system for Methyl-O-tyrosine was procured from the Schultz lab. The procedure followed for the expression of the protein was as follows. 10 and 50 ml cultures were used for expression of CRABPIK106Am. The pET16bCRABPIK106Am plasmid and the pSupJYRS plasmids were co-transformed into BL21 (DE3) cells and single colonies of cells were inoculated into 2XYT medium with ampicillin and chloramphenicol for the preparation of the overnight starter culture. This was used to inoculate the growth culture to start with an OD at 600 nm of 0.2. It took 3 Hrs for the culture to attain an OD of 0.7 at which point the culture was induced with 400 uM IPTG and 1 mM methyl-O-tyrosine was added. The cells were allowed to grow at 37 ^{0}C, 220 rpm for 0.5 Hr. After which the culture was split to two 5 or 25 ml cultures. Both were grown at 30 ^{0}C. One was harvested after 4 Hrs and the other after overnight induction.

Plasmid stability test was performed on LB, LB+ampicillin+chloramphenicol+IPTG, LB+ampicillin+chloramphenicol, LB+IPTG agar plates in case of overnight induction to confirm the presence of both the plasmids during the expression of the protein. It was confirmed that the plasmids were stable during the whole period.

The harvested cultures were lysed with BPER II reagent and the lysates were analyzed by coommassie staining, western blotting using anti-CRABP I antibodies and anti-His tag antibodies.

The harvested cultures were lysed with BPER II reagent and the lysates were analyzed by coommassie staining, western blotting using anti-CRABP I antibodies and anti-His tag antibodies.

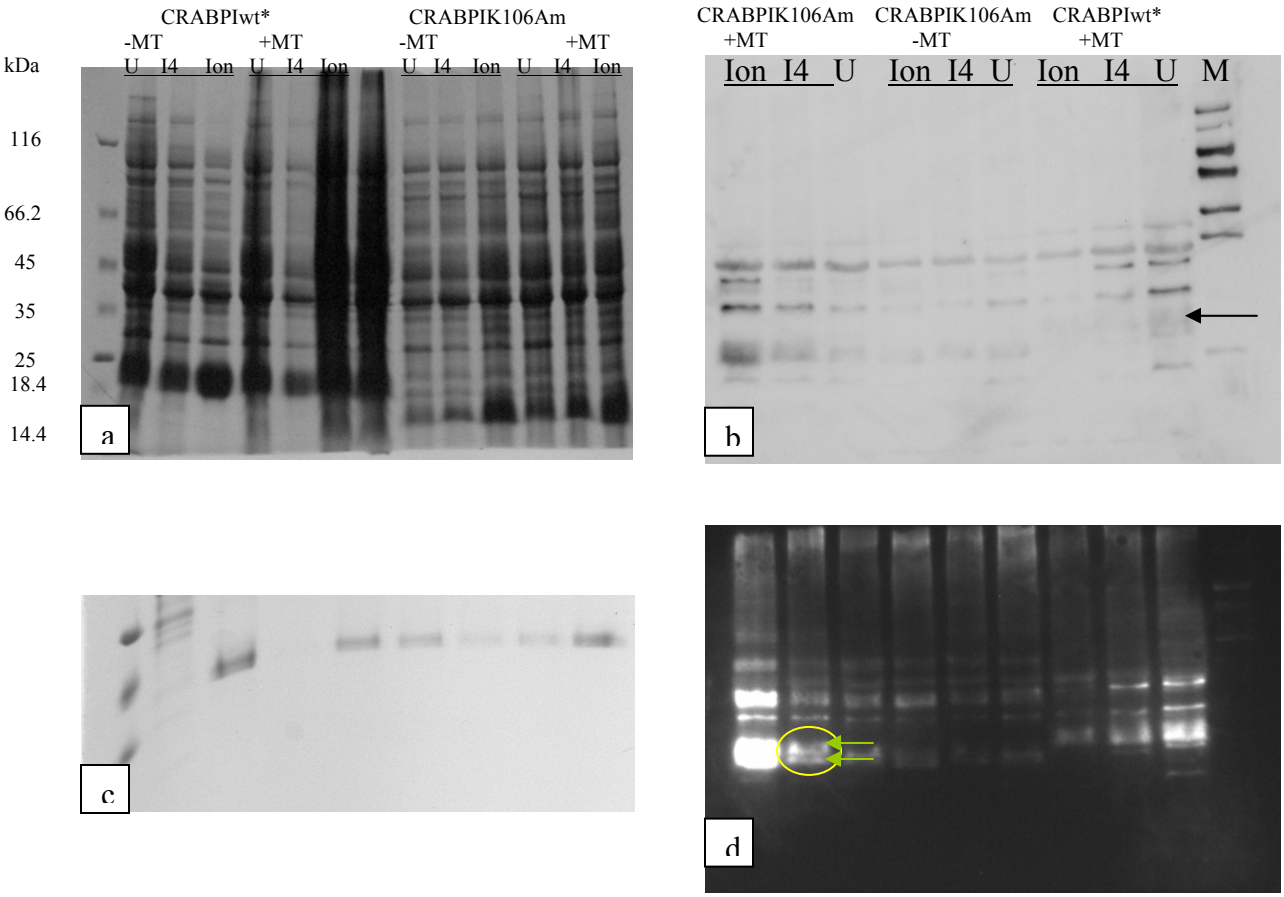


Fig 6.1 Expression of Methyl-O-tyrosine labeled CRABPI. a) Coomassie stained SDS-PAGE gel of the suppression test. b) Western blot using anti-CRABP I antibodies. c) Small scale Ni-NTA purification of the induced overnight fraction of the pET16bCRABPIK106Am+pSUPJYRS +MT culture. d) Western blot as b) with anti-Histag antibodies. MT=Methyl-O-tyrosine, U=Uninduced,I 4=Induced for 4Hrs,Ion=Induced overnight.

In Fig 6.1 a, it is observed that the full length protein is expressed in case of expression from pET16bCRABPIwt*+pSupJYRS both in the absence and in the presence of Methyl-O-tyrosine. This shows that the presence of the amino acid and the specific amino acyl synthetase enzyme does not effect the expression of the wild type protein. The presence of the protein in the uninduced cultures is due to the leakiness of expression. The expression of the protein from the pET16bCRABPIK106Am+pSupJYRS culture on the other hand shows different results. Truncated protein bands can be observed in this case. It was expected that in the presence of the amino acid, full length protein should be expressed with the label and in the absence of the amino acid, only truncated products should be observed. This result is not clear from Fig 6.1 a, but is clear from Fig 6.1 b and

Fig 6.1 d western blots wherein clear bands are observed in case of the samples with the unnatural amino acid while only truncated proteins are observed in samples without the amino acid. In case of the 4 Hr induction sample highlighted in d) two distinct bands are observed, which can be judged as being the full length and the truncated version of the protein. These bands are merged in case of the overnight induction sample and hence appear as one single band. The reason for the presence of multiple bands is not clear and could be a result of degradation and/or oligomerization. It is clear in the western blot results that truncated products are formed to a large extent. This is due to the type of overexpression system used and the mismatch between the protein synthesis machinery and the ribosomal pause at the amber stop codon. A small scale purification of the overnight induced mutant culture with methyl tyrosine (60 µl) was subjected to Ni-NTA purification due to the presence of the N-terminal His tag. The result of which is shown in Fig 6.1 c. The identity of the bands in terms of protein length and presence of incorporated amino acid has to be learnt by mass spectrometry.

6.1. B Expression of the 7-hydroxycoumarinyl-4-ethylglycine labeled CRABPI

Earlier attempts to suppress the K106amber mutation in CRABP I with the fluorescent amino acid was made in a 50 ml culture. The protocol used was similar to the one used for Methyl-O-tyrosine. The cultures were induced overnight and the samples were harvested and lysed in BPER II following which the soluble and insoluble fractions were separated and analyzed. Uninduced cultures were also grown as controls. The fluorescence of the fractions was tested and the identity of the proteins was tested by western blot. Multiple bands were observed (Fig 6.2 b) the smallest being around 15 kDa (Fig 6.2 a) which resembles the truncated version of the protein. A small fraction of the soluble fraction was purified using a Ni-NTA beads and the gel loaded with the samples was exposed to a UV torch (Fig 6.2 c) and a western blot with anti-CRABP I (Fig 6.2 d) antibodies was performed and it was found that a series of truncated products were present. Similar result was observed with Methyl-O-tyrosine labeling and rechecking the sequence of the plasmids lead to the conclusion that these results were due to primer duplication in the sequence due to which two stop codons were inserted. However, the presence of the fluorescent band is good evidence that the method works.

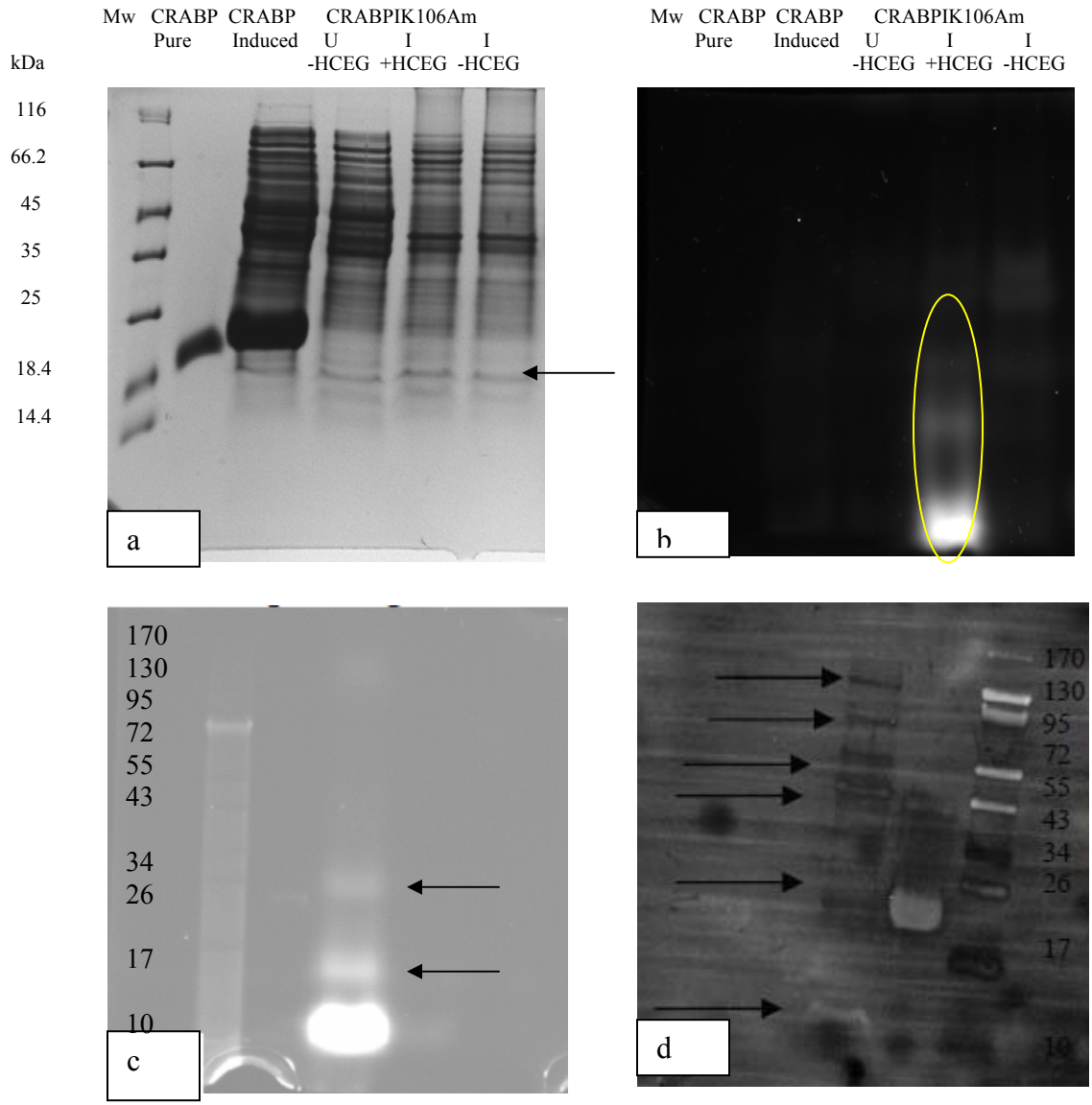


Fig 6.2 Expression of 7-hydroxycoumarinyl-4-ethylglycine labeled CRABP I at K106 position. a) Coomassie stained SDS-PAGE gel. b) UV exposed SDS-PAGE gel prior to coomassie staining. Uninduced cultures and cultures without the addition of the fluorescent amino acid showed no fluorescence. Truncated products observed. The arrow marks show the protein of interest and the truncated and higher molecular weight versions of the same

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

Results of these experiments are promising and give hope for the labeling to occur and further experiments to express the fluorescent amino acid labeled protein with the corrected plasmid sequence will be done consequently. To facilitate the separation of full length labeled protein from truncated products, C-terminal His tagged protein should be used. Choosing an earlier position in the protein sequence would further avoid the synthesis of truncated protein and ease the labeling process. General rules for selection of a site for amber mutation allowing for better suppression of the protein have been learned in this process which will help in choosing sites for incorporation. The pEVOL plasmid system for HCEG incorporation has been requested from the Schultz group. This system will offer a better expression level of the HCEG labeled CRABPI.

In general the performed experiments have paved the way for further studies on *in vivo* incorporation of unnatural amino acids in *E. coli*. Successful incorporation of the fluorescent amino acid into the protein will provide a tool enabling protein dynamics studies by *in vivo* FRET based investigations. These experiments have educated us critical aspects of the nonsense suppression technique which will eventually help in setting up future experiments and exploiting this technique as a tool for exploring the unknown in not only CRABP I folding but also allow for studies in other systems.

BIBLIOGRAPHY

Adamczyk M, Cornwell M, Huff J, Rege S, Rao TVS (1997) Novel 7-hydroxycoumarin based fluorescent labels. Bioorg. *Med. Chem. Lett,7:151985-1988.*

Baslé E, Joubert N, Pucheault M (2010) Protein chemical modification on endogenous amino acids. *Chem Biol,17:213-27.*

Britten RJ, McClure FT (1962) The amino acid pool in Escherichia coli. *292-336.*

Brun MP, Bischoff L, Garbay C (2004) A very short route to enantiomerically pure coumarin-bearing fluorescent amino acids. *Angew Chem Int Ed Engl,43:3432-6.*

Budisa N, Rubini M, Bae JH, Weyher E, Wenger W, Golbik R, Huber R, Moroder L (2002) Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angew Chem Int Ed Engl., 41:4066-9.*

Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, Hu H, Brittain SM, Brinker A, Caldwell J, Bursulaya B, Spraggon G, Brock A, Ryu Y, Uno T, Schultz PG, Geierstanger BH (2008) *In vivo* incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J Am Chem Soc,130:9268-81.*

Chambers I, Frampton J, Goldfarb P, Affara N, McBain W, Harrison PR (1986) The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon *TGA. EMBO J,5:1221-7.*

Chen M, Cai L, Fang Z, Tian H, Gao X, Yao W (2008) Site-specific incorporation of unnatural amino acids into urate oxidase in Escherichia coli. *Protein Sci,17:1827-33.*

Chin JW, Martin AB, King DS, Wang L, Schultz PG (2002) Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc Natl Acad Sci U S A.,99:11020-4.*

Colombano CG, Troccoli OE (1985) Simultaneous evaluation of 7-hydroxycoumarin excited states by synchronous luminescence spectrometry. *Anal. Chem.*, 57: 1907–10.

Deiters, A.; Geierstanger, B.; Schultz, P. G (2005) Site-specific *In Vivo* Labeling of Proteins for NMR Studies. *ChemBioChem 6: 55-58*.

Dougherty DA (2000) Unnatural amino acids as probes of protein structure and function. *Curr Opin Chem Biol.,4:645-52.*

Effraim PR, Wang J, Englander MT, Avins J, Leyh TS, Gonzalez RL Jr, Cornish VW (2009) Natural amino acids do not require their native tRNAs for efficient selection by the ribosome. *Nat Chem Biol,.5:947-53*.

Finkel SR (2010) Read-Through Strategies for Suppression of Nonsense Mutations in Duchenne/Becker Muscular Dystrophy: Aminoglycosides and Ataluren *(PTC124). J Child Neurol,* doi: 10.1177/0883073810371129.

Fleissner MR, Brustad EM, Kálai T, Altenbach C, Cascio D, Peters FB, Hideg K, Peuker S, Schultz PG, Hubbell WL (2009) Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc Natl Acad Sci U S A,106:21637-42.*

Furter R (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Sci,7:419-26.*

Gök E, Oztürk C, Akbay N (2008) Interaction of thyroxine with 7 hydroxycoumarin: a fluorescence quenching study. *J Fluoresc., 18:781-5.*

Gunasekaran K, Hagler AT, Gierasch LM (2004) Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins,54:179-94.*

Hartman MC, Josephson K, Lin CW, Szostak JW (2007) An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS One. 2:e972.*

Hayashi G, Goto Y, Suga H (2010) Ribosome evolution for two artificial amino acids in E. coli. *Chem Biol,17:320-1*.

Hendrickson TL, de Crécy-Lagard V, Schimmel P (2004) Incorporation of nonnatural amino acids into proteins. *Annu Rev Biochem,73:147-76.*

Kiick KL, Weberskirch R, Tirrell DA (2001) Identification of an expanded set of translationally active methionine analogues in Escherichia coli. *FEBS Lett, 502:25-30.*

Köhrer C, Yoo JH, Bennett M, Schaack J, RajBhandary UL (2003) A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. *Chem Biol,10:1095-102.*

Ling J, Reynolds N, Ibba M (2009) Aminoacyl-tRNA synthesis and translational quality control. *Annu Rev Microbiol,63:61-78.*

Link AJ, Tirrell DA (2005) Reassignment of sense codons *in vivo*. *Methods,36:291-8.*

Liu DR, Magliery TJ, Pastrnak M, Schultz PG (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proc Natl Acad Sci U S A,94:10092-7*.

Martin R, Hearn M, Jenny P, Gallant J (1988) Release factor competition is equivalent at strong and weakly suppressed nonsense codons. *Mol Gen Genet,213:144-9.*

Mendes KR, Martinez JA, Kantrowitz ER (2010) Asymmetric allosteric signaling in aspartate transcarbamoylase. *ACS Chem Biol,5:499-506.*

Miller JH (1991) Use of nonsense suppression to generate altered proteins. *Methods Enzymol,208:543-63.*

Mills JH, Lee HS, Liu CC, Wang J, Schultz PG (2009) A genetically encoded direct sensor of antibody-antigen interactions. *Chembiochem, 10:2162-4*.

Nilsson M, Rydén-Aulin M (2003) Glutamine is incorporated at the nonsense codons UAG and UAA in a suppressor-free Escherichia coli strain. *Biochim Biophys Acta,1627:1-6*.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., Schultz, P.G. (1989) A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science, 244:182-8.*

Ohno S, Yokogawa T, Fujii I, Asahara H, Inokuchi H, Nishikawa K (1998) Co-expression of yeast amber suppressor tRNATyr and tyrosyl-tRNA synthetase in Escherichia coli: possibility to expand the genetic code. *J Biochem,.124:1065-8.*

Rosano GL, Ceccarelli EA (2009) Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microb Cell Fact, 24;8:41.*

Ryu Y, Schultz PG (2006) Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat methods: 263-5.*

Turner JM, Graziano J, Spraggon G, Schultz PG (2006) Structural plasticity of an aminoacyl-tRNA synthetase active site. *Proc Natl Acad Sci U S A,103:6483-8*.

Umekage S, Kikuchi Y (2009) *In vivo* circular RNA production using a constitutive promoter for high-level expression. *J Biosci Bioeng,108:354-6.*

Uttamapinant C, White KA, Baruah H, Thompson S, Fernández-Suárez M, Puthenveetil S, Ting AY (2010) A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A., 107:10914-9.*

Wang F, Robbins S, Guo J, Shen W, Schultz PG (2010) Genetic incorporation of unnatural amino acids into proteins in Mycobacterium tuberculosis. *PLoS One, 5:e9354.*

Wang J, Xie J, Schultz PG (2006) A genetically encoded fluorescent amino acid. *J Am Chem Soc,128:8738-9.*

Wang L, Brock A, Schultz PG (2002) Adding L-3-(2-Naphthyl)alanine to the genetic code of E. coli. *J Am Chem Soc,124:1836-7*.

Wang L**,** Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. *Chem Biol,8:883-90*.

Wang L, Schultz PG (2004) Expanding the genetic code. *Angew Chem Int Ed Engl., 44:34-66.*

Young TS, Ahmad I, Yin JA, Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli. J Mol Biol,395:361-74.*

Young TS, Schultz PG (2010) Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem.*, 285: 11039-11044.

Zhang Y, Wang L, Schultz PG, Wilson IA (2005) Crystal structures of apo wild-type M. jannaschii tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein Sci,14:1340-9*.